

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number
WO 02/024736 A1

(51) International Patent Classification: C07K 14/195,
C07H 21/04, C12N 15/52, 15/62

(21) International Application Number: PCT/AU01/01190

(22) International Filing Date:
21 September 2001 (21.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PR 0277 21 September 2000 (21.09.2000) AU
PR 0304 22 September 2000 (22.09.2000) AU
PR 0320 22 September 2000 (22.09.2000) AU

(71) Applicant (for all designated States except US): THE
UNIVERSITY OF QUEENSLAND [AU/AU]; St Lucia,
Queensland 4067 (AU).

(72) Inventor; and

(75) Inventor/Applicant (for US only): BIRCH, Robert
[AU/AU]; 24 Yarimbah Street, Jindalee, Queensland 4074
(AU).

(74) Agents: ARGÆT, Victor, Peter et al.; Davies Collison
Cave, Level 3, 303 Coronation Drive, Milton, Queensland
4064 (AU).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU,
ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:
— with international search report

(48) Date of publication of this corrected version:
7 November 2002

(15) Information about Correction:
see PCT Gazette No. 45/2002 of 7 November 2002, Sec-
tion II

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH ANTIBIOTIC BIOSYNTHESIS AND USES
THEREFOR

(57) Abstract: The present invention discloses polyketides and the polyketide synthases and ancillary enzymes that are capable of producing such compounds. More particularly, the present invention discloses polynucleotides and polypeptides associated with (i) a novel polyketide synthase linked to a non-ribosomal peptide synthetase involved in the biosynthesis of albidins, (ii) a novel phosphopantetheinyl transferase for activating enzymes, particularly polyketide synthases and/or non-ribosomal peptide synthetases, associated with the biosynthesis of albidins, and (iii) a novel methyltransferase for methylating precursors of albidins and/or intermediates related to albidin biosynthesis. The present invention also discloses methods of using the aforementioned polynucleotides and polypeptides for activating polyketide synthases and/or non-ribosomal peptide synthetases, for methylating precursors of albidins or their analogues and/or intermediates involved in the biosynthesis of albidins or analogues thereof and for enhancing the level and/or functional activity of albidins or their analogues. Also disclosed are methods of using the polynucleotides and polypeptides of the invention for the biosynthesis of albidins or their analogues.

WO 02/024736 A1

POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH ANTIBIOTIC BIOSYNTHESIS AND USES THEREFOR

FIELD OF THE INVENTION

THIS INVENTION relates generally to antibiotic biosynthesis. More particularly, the present invention relates to polyketides and the polyketide synthases and ancillary enzymes that are capable of producing such compounds. Even more particularly, the present invention relates to a polyketide synthase linked to a non-ribosomal peptide synthetase involved in the biosynthesis of albicidins, to a phosphopantetheinyl transferase for activating enzymes, particularly polyketide synthases and/or non-ribosomal peptide synthetases, associated with the biosynthesis of albicidins, and to a methyltransferase for methylating precursors of albicidins and/or intermediates related to albicidin biosynthesis. The present invention also relates to biologically active fragments of the aforementioned polypeptides and to variants and derivatives of these molecules. Further, the invention relates to polynucleotides encoding the said polypeptides, including the *xabA*, *xabB* and *xabC* genes of *Xanthomonas albilineans*, to polynucleotides encoding the said fragments, variants or derivatives, to vectors comprising the said polynucleotides and to host cells containing such vectors. The invention also relates to a transcriptional control element for modulating the expression of polynucleotides including, for example, the *xabB* gene and/or the *xabC* gene of *Xanthomonas albilineans*, or variants thereof. The invention also features methods of using the polynucleotides, polypeptides, fragments, variants, derivatives and vectors for activating polyketide synthases and/or non-ribosomal peptide synthetases, for methylating precursors of albicidins or their analogues and/or intermediates involved in the biosynthesis of albicidins or their analogues and for enhancing the level and/or functional activity of albicidins or their analogues. The invention also encompasses methods of using the aforesaid polynucleotides, polypeptides, fragments, variants and derivatives for the biosynthesis of albicidins or analogues thereof.

Bibliographic details of various publications referred to by author in this specification are collected at the end of the description.

- 2 -

BACKGROUND OF THE INVENTION

Polyketides represent a large structurally diverse group of compounds synthesised from 2-carbon units through a series of condensations and subsequent modifications. They possess a broad range of biological activities including antibiotic and pharmacological properties. For example, polyketides are represented by antibiotics such as tetracyclines, erythromycins, immunosuppressants such as FK506, FK520 and rapamycin, anticancer agents such as daunomycin and veterinary products such as monensin and avermectin.

Considering the difficulty in producing polyketide compounds by conventional chemical methodologies, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds. In this regard, reference may be made to PCT publication Nos. WO 93/13663; WO 95/08548; WO 96/40968; WO 97/02358; and WO 98/27203; U.S. Pat. Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; and 5,712,146; Fu *et al.* (1994, *Biochemistry* 33: 9321-9326); McDaniel *et al.* (1993, *Science* 262: 1546-1550); and Rohr (1995, *Angew. Chem. Int. Ed. Engl.* 34(8): 881-888).

Polyketides are synthesised in nature by polyketide synthases (PKS). These enzymes, which are actually complexes of multiple enzyme activities, are in some ways similar to, but in other ways different from, the synthases that catalyse condensation of 2-carbon units in the biosynthesis of fatty acids. Specifically, PKS enzymes catalyse the biosynthesis of polyketides through repeated (decarboxylative) Claisen condensations between acylthioesters (*e.g.*, acetyl, propionyl, malonyl or methylmalonyl). Following each condensation, they introduce structural variability into the product by catalysing all, part, or none of a reductive cycle comprising a ketoreduction, dehydration, and enoylreduction on the β -keto group of the growing polyketide chain. PKS enzymes incorporate enormous structural diversity into their products, in addition to varying the condensation cycle, by controlling choice of primer, extender units, and the overall chain length and, particularly in the case of aromatic polyketides, regiospecific cyclisation of the nascent polyketide chain. After the carbon chain has grown to a length characteristic of each specific product, it is released from the synthase by thiolysis or acyltransfer. Thus, the PKS complexes consist of families of enzymes which work together to produce a given polyketide. It is the choice of chain-building units, controlled variation in chain length, and the reductive cycle,

- 3 -

genetically programmed into each PKS, that contributes to the variation seen among naturally occurring polyketides.

Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis of the polyketide synthesised. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes. These classifications are well known and reference may be made, for example, to Hopwood and Khosla (1992).

The Type I or modular PKS enzymes typically catalyse the biosynthesis of complex polyketides such as erythromycin and avermectin. These modular enzymes include assemblies of several large multifunctional proteins carrying, between them, a set of separate active sites for each step of carbon chain assembly and modification (Cortes *et al.*, 1990; Donadio *et al.*, 1991; MacNeil *et al.*, 1992). Accordingly, modular PKS complexes can be viewed as biochemical assembly lines, composed of a series of catalytic domains involved in sequential assembly and modification of acyl groups on the growing polyketide chain (Cane *et al.*, 1998; Keating and Walsh, 1999). The catalytic domains are arranged in "modules", punctuated by acyl carrier protein (ACP) domains that tether the nascent polyketide while it undergoes the catalytic modifications programmed in the associated module. For each polyketide there is an initiation module, a series of elongation modules that define the length and structure of the polyketide chain, and a termination module to release the product from the final tether. The initiation module typically comprises an acyl transferase (AT) domain that couples the initial acyl group from an acyl-CoA substrate to the phosphopantetheinyl tether of the first ACP domain. Each elongation module typically comprises a ketosynthase (KS), an AT and an ACP. The KS removes the growing polyketide unit from the upstream ACP and couples it to the next acyl group in the chain, which has already been selected and loaded by the AT onto the ACP in the same module. Other catalytic domains (*eg.* a ketoacyl reductase (KR), and dehydratase (DH)) within an elongation module can modify the newly elongated polyketide before it is transferred to the next module in the biochemical assembly line. A thioesterase (TE) domain in the termination module accomplishes release of the assembled polyketide from the last ACP in the series (Cane *et al.*, 1998; Keating and Walsh, 1999).

Biosynthesis of a polyketide can involve the sequential action of several PKS proteins, each with one to six elongation modules (MacNeil *et al.*, 1992; Apricio *et al.*, 1996). There are variations on the modular PKS design, including participation by some loading domains across modules or in *trans* from separate proteins (Keating and Walsh, 5 1999), and several examples of hybrid PKS/NRPS proteins (Albertini *et al.*, 1995; Gehring *et al.*, 1998; Duitman *et al.*, 1999; Paitan *et al.*, 1999). Subsequent modification of the polyketide by dedicated tailoring enzymes is generally required to complete the biologically active product (Hopwood, 1997). Other biologically active compounds including antibiotics comprise polypeptides assembled by non-ribosomal peptide 10 synthetases (NRPSs). NRPSs typically show a modular architecture and tethered biosynthetic strategy analogous to PKSs (Cane *et al.*, 1998; Keating and Walsh, 1999). In NRPSs a condensation (C) domain removes the growing peptide unit from the upstream PCP domain and couples it to the next amino acid group in the chain, which has already been selected and loaded by an adenylation (A) domain onto the PCP in the same module 15 (Marahiel *et al.*, 1997; Stachelhaus *et al.*, 1998). Other catalytic domains (*e.g.*, epimerase or N-methyltransferase) within an elongation module can modify the newly elongated polypeptide before it is transferred to the next module in the biochemical assembly line (Marahiel *et al.*, 1997).

Many phytopathogenic bacteria and fungi secrete toxins with phytotoxic activity 20 and a broad spectrum of antimicrobial properties (Guenzi *et al.*, 1998). Albicidin phytotoxins are polyketides produced by *Xanthomonas albilineans*, which are key pathogenicity factors in the development of leaf scald, one of the most devastating diseases of sugarcane (*Saccharum*, interspecific hybrids) (Ricaud and Ryan, 1989; Zhang and Birch, 1997; Zhang *et al.*, 1999). Albicidins selectively block prokaryote DNA replication and cause 25 the characteristic chlorotic symptoms of leaf scald disease by blocking chloroplast development (Birch and Patil, 1983; 1985b; 1987a; 1987b). Because albidins are rapidly bactericidal at nanomolar concentrations against a broad range of Gram-positive and Gram-negative bacteria, they are also of interest as potential clinical antibiotics (Birch and Patil, 1985a).

30 The major antimicrobial component of the family of albidins produced in culture by *X. albilineans* has been partially characterised as a low M_r compound with several aromatic rings (Birch and Patil, 1985a). Low yields have slowed studies into the

- 5 -

chemical structure of albicidin, its application as a tool to study prokaryote DNA replication, and its development as a clinical antibiotic (Zhang *et al.*, 1998). Genetic analysis of albicidin biosynthesis is likely to indicate approaches to increase yields, probable structural features, and opportunities for engineering novel antibiotics in this

5 family.

SUMMARY OF THE INVENTION

The present invention arises in part from the identification and characterisation of several *X. albilineans* genes associated with albicidin biosynthesis. In particular, the present inventor has isolated a novel *X. albilineans* gene (*xabB*), which encodes a large protein (predicted Mr 525,695), with a modular architecture indicative of a multifunctional PKS linked to a non-ribosomal peptide synthetase (NRPS). At 4801 amino acids in length, the product of *xabB* (XabB) is the largest reported PKS-NRPS. Twelve catalytic domains in this multifunctional enzyme are arranged in the order N-terminus-acyl-CoA ligase (AL)-acyl carrier protein (ACP)- β -ketoacyl synthase (KS)- β -ketoacyl reductase (KR)-ACP-ACP-KS-peptidyl carrier protein (PCP)-condensation domain (C)-adenylation domain (A)-PCP-C. The modular architecture of XabB indicates likely steps in albicidin biosynthesis, and approaches to enhance antibiotic yield. The novel pattern of domains, in comparison with known PKS-NRPS enzymes for antibiotic production, also contributes to the knowledge base for rational design of enzymes producing novel antibiotics. The present inventor has found that XabB is required for the production of albicidins and that enhanced expression of *xabB* leads to increased levels and/or functional activities of albicidin antibiotics.

A gene (*xabC*) encoding a novel O-methyltransferase has also been isolated, which methylates albicidin precursors and/or intermediates involved in albicidin biosynthesis. Surprisingly, enhanced expression of *xabC* has been found to increase the levels and/or functional activities of albicidin antibiotics.

The present inventor has also isolated a gene (*xabA*) encoding a phosphopantetheinyl transferase (PPTase), which is required for post-translational activation of synthetases in the albicidin biosynthetic pathway. In this regard, it is known that inefficient phosphopantetheinylation has limited the activity of other antibiotic synthetases overexpressed in heterologous species (Walsh *et al.*, 1997). Accordingly, the isolated *xabA* gene, together with its target in the albicidin biosynthetic pathway (e.g., *xabB*), provide the means to engineer high level co-expression of the albicidin synthetase and its activating PPTase to obtain albicidins in higher yields, and ultimately to manipulate the elements of the albicidin biosynthetic machinery, by mutagenesis or by other means, to produce desired structural variants of this novel antibiotic class.

- 7 -

The above genes, in whole or in part, together with their variants and derivatives, are useful *inter alia* for modulating the level and/or functional activity of albicidins, for expressing PKS enzymes in recombinant host cells, for producing polyketides including albicidins and their analogues and for combinatorial biosynthesis, as described hereinafter.

5 Accordingly, one aspect of the present invention contemplates an isolated polypeptide encoding at least a portion of an albicidin PKS-NRPS (XabB) or its variants or derivatives. In one embodiment of this type, the invention provides an isolated polypeptide comprising at least one domain selected from the group consisting of:

10 (a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or more of SEQ ID NO: 6 and 8, or variants thereof.

 (b) a β -ketoacyl synthase (KS) domain comprising a sequence set forth in any one or more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;

 (c) a β -ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID NO: 22, or variants thereof;

15 (d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;

 (e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof;

20 (f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 50 and 52, or variants thereof; and

 (g) a condensation (C) domain comprising a sequence set forth in any one or more of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

25 Preferably, the AL domain comprises each of the sequences set forth in SEQ ID NO: 6 and 8, or variants thereof.

 In one embodiment, the KS domain preferably comprises each of the sequences set forth in SEQ ID NO: 10, 12 and 14, or variants thereof. In an alternate embodiment, the KS domain preferably comprises each of the sequences set forth in SEQ ID NO: 16, 18 and 20, or variants thereof.

30 Preferably, the A domain comprises each of the sequences set forth in SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.

- 8 -

In one embodiment, the C domain preferably comprises each of the sequences set forth in SEQ ID NO: 54, 56, 58, 60, 62, 64 and 66, or variants thereof. In an alternate embodiment, the C domain preferably comprises each of the sequences set forth in SEQ ID NO: 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

- 5 In another embodiment, the invention provides an isolated polypeptide comprising at least a biologically active fragment or portion of the sequence set forth in SEQ ID NO: 2, or a variant or derivative thereof.

Suitably, the biologically active fragment is at least 6 amino acids in length.

- 10 In a preferred embodiment, the domains broadly described above are arranged in an N- to C-terminal direction as follows: AL-ACP-KS-KR-ACP-ACP-KS-PCP-C-A-PCP-C.

Suitably, the biologically active fragment comprises at least one domain selected from the group consisting of the AL domain, the KS domain, the KR domain, the ACP domain, the A domain, the PCP domain and the C domain as broadly described above.

- 15 Suitably, the variant has at least 60%, preferably at least 70%, more preferably at least 80%, more preferably at least 90% and still more preferably at least 95% sequence identity to the sequence set forth in SEQ ID NO: 2.

- 20 Preferably, the variant comprises at least one sequence selected from the group consisting of SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variant thereof. In this regard, the variant preferably has at least 70%, preferably at least 80%, more preferably at least 90%, and still more preferably at least 95% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 25 70, 72, 74, 76, 78 and 80.

In another aspect, the present invention contemplates an isolated polypeptide encoding at least a portion of a PPTase (XabA) associated with albicidin biosynthesis or its variants or derivatives. In one embodiment of this type, the invention provides an isolated

- 9 -

polypeptide comprising at least biologically active fragment or portion of the sequence set forth in SEQ ID NO: 83, or a variant or derivative thereof.

Suitably, the biologically active fragment comprises at least one, and preferably both, of the consensus PPTase sequence motifs set forth in SEQ ID NO: 89 and 93, or
5 variant thereof. Preferably, the biologically active fragment comprises the intervening sequence between the said consensus PPTase sequence motifs, which intervening sequence comprises the sequence set forth in SEQ ID NO: 91, or variant thereof.

Preferably, the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof.

10 Suitably, the variant has at least 60%, preferably at least 70%, more preferably at least 80%, more preferably at least 90% and still more preferably at least 95% sequence identity to the sequence set forth in SEQ ID NO: 83.

Preferably, the variant comprises at least one sequence selected from the group consisting of SEQ ID NO: 87, 89, 91 and 93, or variant thereof. In this regard, the variant
15 preferably has at least 70%, preferably at least 80%, more preferably at least 90%, and still more preferably at least 95% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 87, 89, 91 or 93.

In yet another aspect, the present invention contemplates an isolated polypeptide encoding at least a portion of a methyltransferase (XabC) associated with albicidin
20 biosynthesis or its variants or derivatives. In one embodiment of this type, the invention provides an isolated polypeptide comprising at least biologically active fragment or portion of the sequence set forth in SEQ ID NO: 95, or a variant or derivative thereof.

Suitably, the biologically active fragment comprises at least one, and preferably all, of the consensus methyltransferase sequence motifs set forth in SEQ ID NO: 99, 101
25 and 103, or variant thereof.

Preferably, the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 105, or variant thereof. In a preferred embodiment, the biologically active fragment comprises a contiguous

- 10 -

sequence of amino acids contained within the sequence set forth in SEQ ID NO: 107, or variant thereof.

Suitably, the variant has at least 60%, preferably at least 70%, more preferably at least 80%, more preferably at least 90% and still more preferably at least 95% sequence identity to the sequence set forth in SEQ ID NO: 95.

Preferably, the variant has at least 70%, preferably at least 80%, more preferably at least 90%, and still more preferably at least 95% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 99, 101 and 103.

In still yet another aspect, the invention contemplates an isolated polynucleotide encoding at least a portion of an albicidin PKS-NRPS (XabB) or its variants or derivatives, as broadly described above. Preferably, the polynucleotide comprises the sequence set forth in any one of SEQ ID NO: 1 and 3, or a biologically active fragment thereof, or a polynucleotide variant of these.

Suitably, the biologically active fragment is at least 18 nucleotides in length.

The polynucleotide preferably encodes at least one domain selected from the group consisting of the AL domain, the KS domain, the KR domain, the ACP domain, the A domain, the PCP domain and the C domain as broadly described above.

Suitably, the AL domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 5 and 7, or variants thereof. Preferably, the AL domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 5 and 7, or variants thereof.

The KS domain is preferably encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 9, 11, 13, 15, 17 and 19, or variants thereof. In one embodiment, the KS domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 9, 11 and 13, or variants thereof. In an alternate embodiment, the KS domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 15, 17 and 19, or variants thereof.

- 11 -

Preferably, the KR domain is encoded by a nucleotide sequence set forth in SEQ ID NO: 21, or variant thereof.

Suitably, the ACP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 23, 25 and 27, or variants thereof.

5 The A domain is preferably encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof. In a preferred embodiment, the A domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof.

10 Suitably, the PCP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 49 and 51, or variants thereof.

15 Preferably, the C domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79, or variants thereof. In one embodiment, the C domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 53, 55, 57, 59, 61, 63 and 65, or variants thereof. In an alternate embodiment, the C domain is preferably encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 67, 69, 71, 73, 75, 77 and 79, or variants thereof.

20 In one embodiment, the polynucleotide variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 1 or 3.

25 In another embodiment, the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 1 or 3 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

Preferably, the polynucleotide variant comprises a nucleotide sequence encoding at least one domain selected from the group consisting of the AL domain, the KS domain, the KR domain, the ACP domain, the A domain, the PCP domain and the C domain as broadly described above.

- 12 -

In one embodiment, the nucleotide sequence variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the sequences set forth in SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79.

In another embodiment, the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

In a further aspect, the invention contemplates an isolated polynucleotide encoding at least a portion of a PPTase (XabA) associated with albicidin biosynthesis or its variants or derivatives. an isolated polynucleotide encoding a polypeptide, fragment, variant or derivative as broadly described above. Preferably, the polynucleotide comprises the sequence set forth in any one of SEQ ID NO: 82 and 84, or a biologically active fragment thereof, or a polynucleotide variant of these.

Alternatively, the polynucleotide comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 86, or variant thereof.

In one embodiment, the polynucleotide variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 82, 84 and 86.

In another embodiment, the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 82, 84 and 86 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

Preferably, the polynucleotide variant comprises a nucleotide sequence encoding at least one PPTase sequence motif selected from SEQ ID NO: 89 and 93, or variant thereof. Suitably, the polynucleotide variant comprises a nucleotide sequence encoding the

- 13 -

intervening sequence between the said consensus PPTase sequence motifs, said nucleotide sequence comprising the sequence set forth in SEQ ID NO: 91.

The polynucleotide variant suitably comprises a nucleotide sequence encoding a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof. In this instance, the contiguous sequence is preferably encoded
5 by the sequence set forth in SEQ ID NO: 86, or nucleotide sequence variant thereof

Suitably, the PPTase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 88 and 92, or nucleotide sequence variant thereof.

10 Preferably, the said intervening sequence is encoded by the nucleotide sequence set forth in SEQ ID NO: 90, or nucleotide sequence variant thereof.

In one embodiment, the nucleotide sequence variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the sequences set forth in SEQ ID NO: 86, 88, 90 and 92.

15 In another embodiment, the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 86, 88, 90 and 92 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

In yet a further aspect, the invention contemplates an isolated polynucleotide
20 encoding at least a portion of a methyltransferase (XabC) associated with albicidin biosynthesis or its variants or derivatives. Preferably, the polynucleotide comprises the sequence set forth in any one of SEQ ID NO: 94 and 96, or a biologically active fragment thereof, or a polynucleotide variant of these.

Alternatively the polynucleotide comprises a contiguous sequence of nucleotides
25 contained within the sequence set forth in SEQ ID NO: 104, or variant thereof. In one embodiment, this polynucleotide preferably comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 106, or variant thereof

- 14 -

In one embodiment, the polynucleotide variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 94, 96, 104 and 106.

5 In another embodiment, the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 94, 96, 104 and 106 under at least low stringency conditions, preferably under at least medium stringency conditions, and more preferably under high stringency conditions.

10 Preferably, the polynucleotide variant comprises a nucleotide sequence encoding a methyltransferase sequence motif selected from any one or more of SEQ ID NO: 99, 101 and 103, or variant thereof.

Suitably, the methyltransferase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 98, 100 and 102, or nucleotide sequence variant thereof.

15 In one embodiment, the nucleotide sequence variant has at least 60%, preferably at least 70%, more preferably at least 80%, and still more preferably at least 90% sequence identity to any one of the sequences set forth in SEQ ID NO: 98, 100 and 102.

In another embodiment, the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 98, 100 and 102 under at least low stringency conditions, preferably under at least medium stringency conditions, and more
20 preferably under high stringency conditions.

In still a further aspect, the invention features an expression vector comprising a polynucleotide as broadly described above wherein the polynucleotide is operably linked to a regulatory polynucleotide.

25 In another aspect, the invention provides a host cell containing a said expression vector.

Suitably, the host cell is a bacterium or other prokaryote.

In yet another aspect, the invention is directed to a multiplicity of cell colonies, constituting a library of colonies, wherein each colony of the library contains an expression

- 15 -

vector for the production of a polypeptide, fragment, variant or derivative as broadly described above.

The invention also features a method of producing a recombinant polypeptide, fragment, variant or derivative as broadly described above, comprising:

- 5 – culturing a host cell containing an expression vector as broadly described above such that said recombinant polypeptide, fragment, variant or derivative is expressed from said polynucleotide; and
- isolating the said recombinant polypeptide, fragment, variant or derivative.

In another aspect, the invention provides a method of producing a biologically active fragment of a polypeptide as broadly described above, comprising:

- 10 – detecting an activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 2, wherein said activity is selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity and condensation activity;
- 15 or
- detecting PPTase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 83; or
- detecting methyltransferase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 95;
- 20 wherein detection of said activity is indicative of said fragment being a biologically active fragment.

In a further aspect, the invention provides a method of producing a biologically active fragment as broadly described above, comprising:

- 25 – introducing a polynucleotide from which a fragment of a polypeptide as broadly described above can be produced into a cell; and
- detecting an activity selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity and condensation activity; or
- detecting PPTase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 83; or
- 30

- 16 -

— detecting methyltransferase activity associated with a fragment of the polypeptide set forth in SEQ ID NO: 95;

wherein detection of said activity is indicative of said fragment being a biologically active fragment.

5 In yet a further aspect, the invention provides a method of producing a variant of a polypeptide as broadly described above (parent polypeptide), or a biologically active fragment thereof, comprising:

— producing a modified polypeptide whose sequence is distinguished from the parent polypeptide or the biologically active fragment by substitution, deletion or
10 addition of at least one amino acid; and

— detecting an activity associated with the modified polypeptide, wherein said activity is selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and
15 methyltransferase activity, wherein detection of said activity is indicative of said modified polypeptide being a variant.

In a further aspect, the invention contemplates a method of producing a variant of a parent polypeptide as broadly described above, or biologically active fragment thereof, comprising:

20 — producing a polynucleotide from which a modified polypeptide as described above can be produced;

— introducing said polynucleotide into a cell; and

— detecting an activity associated with the modified polypeptide, wherein said activity is selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl
25 synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity, wherein detection of said activity is indicative of said modified polypeptide being a variant..

30 In yet another aspect, the invention extends to a method of screening for an agent that modulates the expression of a gene or variant thereof or the level and/or functional activity of an expression product of said gene or variant thereof, wherein said gene is

- 17 -

selected from *xabB*, *xabA*, or *xabC*, or a gene belonging to the same regulatory or biosynthetic pathway as *xabB*, *xabA*, or *xabC*, said method comprising:

- 5 - contacting a preparation comprising a polypeptide encoded by said gene, or biologically active fragment of said polypeptide, or variant or derivative of these, or a genetic sequence (*e.g.*, a transcriptional control element) that modulates the expression of said gene or variant thereof, with a test agent; and
- detecting a change in the level and/or functional activity of said polypeptide or biologically active fragment thereof, or variant or derivative, or of a product expressed from said genetic sequence.

10 The transcriptional control element preferably comprises the sequence set forth in SEQ ID NO: 81 or complement thereof.

The invention, in another aspect, also provides a method for enhancing the level and/or functional activity of an albicidin, said method comprising:

- 15 - introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of an albicidin PKS-NRPS or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of an albicidin PKS-NRPS or variant or derivative thereof can be translated;
- 20 - and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin.

Preferably, the method further comprises introducing into said host cell a vector from which a PPTase can be translated. Suitably, the PPTase is selected from EntD or XabA.

25 Preferably, the method further comprises introducing into said host cell a vector from which a methyltransferase, more preferably and *O*-methyltransferase, and even more preferably an *S*-adenosylmethionine *O*-methyltransferase can be translated.

30 According to another aspect of the invention, there is provided a method for enhancing the level and/or functional activity of an albicidin, said method comprising contacting a precursor of said albicidin or an intermediate involved in the biosynthesis of said albicidin with at least a portion of an albicidin PKS-NRPS, or variant or derivative

- 18 -

thereof, as broadly described above, for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin.

Preferably, the method further comprises contacting a precursor of said albicidin or an intermediate involved in the biosynthesis of said albicidin with a PPTase.

5 Preferably, the method further comprises contacting a precursor of said albicidin or an intermediate involved in the biosynthesis of said albicidin with a methyltransferase, more preferably an *O*-methyltransferase, and even more preferably an *S*-adenosylmethionine *O*-methyltransferase.

10 In another aspect, the invention provides a method of identifying a PPTase for enhancing the level and/or functional activity of an albicidin, said method comprising introducing into an albicidin-deficient strain of *X. albilineans* which lacks *xabA* a vector comprising a polynucleotide encoding a test PPTase, wherein said polynucleotide is operably linked to a regulatory polynucleotide, and detecting production of albicidin.

Suitably, the strain is LS156 described herein.

15 Preferably, the PPTase is EntD.

The invention, in another aspect, also provides a method for enhancing the level and/or functional activity of an albicidin, said method comprising:

- 20 – introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of a PPTase associated with albicidin biosynthesis or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of a PPTase associated with albicidin biosynthesis or variant or derivative thereof can be translated;
- 25 – and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin

In yet another aspect, the invention provides a method for enhancing the level and/or functional activity of an albicidin, said method comprising:

- introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of a methyltransferase associated

- 19 -

with albicidin biosynthesis or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of a methyltransferase associated with albicidin biosynthesis or variant or derivative thereof can be translated;

- 5 - and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin

In another aspect, the invention resides in an antigen-binding molecule that is immuno-interactive with a polypeptide, fragment, variant or derivative as broadly described above.

- 10 In yet another aspect, the invention provides a method to prepare a polynucleotide encoding a modified PKS, comprising using an albicidin PKS-NRPS encoding nucleotide sequence as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, deletion, insertion, or replacement.

- 15 In still yet another aspect, the invention contemplates a method for producing polyketides, comprising expressing the modified albicidin PKS encoding nucleotide sequence as broadly described in a suitable host cell to thereby produce a polyketide different from that produced by the albicidin PKS-NRPS.

- 20 Another aspect of the invention contemplates the insertion of portions of the albicidin PKS-NRPS coding sequence into other PKS coding sequences to modify the products thereof.

- 25 In a further aspect, the invention encompasses use of the polypeptide, fragment, variant or derivative as broadly described above, or the polynucleotide or vector as broadly described above, or the modulatory agent as broadly described above for producing secondary metabolites, preferably albicidins.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation showing a physical and functional map of part of the albicidin biosynthetic gene cluster. (A). Partial physical map of the Tn5 insertion locus in LS157 genomic DNA. Restriction enzymes used: C, *Cla*I; E, *Eco*RI; S, *Spe*I; N, *Not*I; and B, *Bam*HI. (B). Probes used to recover clone pXABB: Probe 1, 1.4-kb *Eco*RI-*Not*I fragment digested from pBC157; and probe 2, 0.9-kb PCR product amplified from Xa13 genomic DNA using primers complementary to sequences flanking the Tn5 insertion in LS157. (C). Clones and subclones used for sequencing, and described in Table 1. (D). The transcription directions of three putative ORFs in 16.5-kb *Eco*RI fragment are indicated by arrows. (E). Organisation of *X. albilineans* XabB constructed by comparison with known protein sequences. The unshaded box indicates PKS region, and the shaded box indicates NRPS region. Relative positions of potential catalytic domains or active sites are indicated by: AL, acyl-CoA ligase; ACP, acyl carrier protein; KS, β -ketoacyl synthase; KR, β -ketoacyl reductase; PCP, peptidyl carrier protein; C, condensation; A, adenylation. Horizontal bars indicate proposed biosynthetic modules.

Figure 2 is a diagrammatic representation presenting the sequence of the region upstream from *xabB*. The nucleotide sequence is numbered according to the 16511-bp sequence in GenBank accession no. AF239749. The putative -35 and -10 promoter sequences of *xabB* and the divergent gene *xatA* are underlined, as are ribosome-binding sequences. The transcriptional directions of *xabB* and *xatA* are indicated by arrows. Translational start codons are indicated by boldface type. Primers P1F1 and P1R are shaded.

Figure 3 is a diagrammatic representation showing the alignment of *X. albilineans* XabB enzymatic domains with those of PKSs and FASs from other organisms. Identical amino acids are indicated by boldface type. Stars and overlines identify conserved amino acids at catalytic sites. Xal-XabB, *X. albilineans* XabB for biosynthesis of albicidin (this study); Hin-LCFA, *Haemophilus influenza* long-chain fatty acid-CoA ligase (P46450); Bsu-PksJ, *B. subtilis* polyketide synthase J (P40806); Bsu-MycA, *B. subtilis* MycA for biosynthesis of mycosubtilin (AF184956); Pcr-ComL2, *Petroselinum crispum* 4-coumarate-CoA ligase 2 (P14913); Sma-FkbB, *S. sp.* MA6548 FkbB for biosynthesis of FK506 (AF082099); Ame-RifA, *Amycolatopsis mediterranei* RifA for biosynthesis of

rifamycin B (AF040570); Shy-RapA, *S. hygrosopicus* RapA for biosynthesis of rapamycin (X86780); Mxa-Ta1, *M. xanthus* Ta1 for biosynthesis of TA (AJ006977); Ser-EryA1 and EryA3, *S. erythraea* EryA modules for biosynthesis of erythromycin (M63676, M63677); Che-PKS1, *Cochliobolus heterostrophus* PKS1 for biosynthesis of T-toxin
 5 (U68040); Bsu-PksM, *B. subtilis* PKS for a polyketide synthase (O31781); Mtu-PpsA, *M. tuberculosis* PKS for a polyketide synthase (G3261605); Mtu-MAS, *M. tuberculosis* MAS for biosynthesis of mycocerosic acid (M95808); Chick-FAS, chicken fatty acid synthase (M22987); Rat-FAS, rat fatty acid synthase (X14175).

Figure 4 is a graphical representation showing albicidin production by wild-type
 10 *X. albilineans* LS155 (▲), complemented Tox⁻ mutant strain LS157 pLXABB1 (○), complemented Tox⁻ mutant strain LS157 pLXABB2 (●), LS157 (■), and LS157 pLAFR3 (+). Albicidin concentrations in culture supernatants were quantified based on inhibition zone width in a microbial bioassay (means +/- standard errors from 5 replicates).

Figure 5 is a graphical representation showing the relationship between growth
 15 (■), albicidin production (○), and GUS activity (▲) in *X. albilineans* LS155 pRG960p1 (A) and in LS155 pRG960p2 (B). Relative activity (means +/- standard errors from 2 replicates): 100% growth, OD₅₅₀ = 1.43; 100% albicidin production = 268.5 units/ml; 100% GUS activity = 119 units/mg of protein (one unit equals 1 pmol of methylumbelliferone formed per min.). Locations and sizes of inserts on pRG960p1 and
 20 pRG960p2 are indicated in Figure 2 and Table 1. GUS, β -glucuronidase.

Figure 6 is a schematic representation showing the organisation of five known PKS-NRPS enzymes. *X. albilineans* XabB, encoded by *xabB* for albicidin biosynthesis (this study); *B. subtilis* MycA for mycosubtilin biosynthesis (Duitman *et al.*, 1999); *Yersinia pestis* HMWP1 for yersiniabactin biosynthesis (Gehring *et al.*, 1998); *M. xanthus*
 25 partial gene product Ta1 for TA biosynthesis (Paitan *et al.*, 1999); *B. subtilis* PksorFX6 for unknown function (Albertini *et al.*, 1995). Unshaded boxes indicate PKS regions, grey boxes indicate NRPS regions, and dark boxes indicate amino transferase (AMT) or methyltransferase (MT). Vertical bars follow the carrier domains at the end of each biosynthetic "module".

30 Figure 7 is a diagrammatic representation showing a dendrogram (GCG) analysis of adenylation domains of XabB and its homologous peptide synthetases. Peptide

- 22 -

synthetases, including various modules of the same multienzyme complex, are as follows: GrsA and GrsB, gramicidin synthetase A and B, respectively, from *B. subtilis* (X15577, X61658); BacA, BacB, and BacC, bacitracin synthetase A, B, and C, respectively, from *B. licheniformis* (AF007865); SnbC and SnbDE, pristinamycin I synthetase C and DE, respectively, from *S. pristinaespiralis* (X98690, Y11547); FkbP, FK506 synthetase FkbP from *S. sp.* MA6548 (AF082100); TycA, TycB, and TycC, tyrocidine synthetase A, B, and C, respectively, from *B. brevis* (AF004835); SyrE, syringomycin synthetase E1 from *Pseudomonas syringae* pv. *syringae* (AF047828); EntF, enterobactin synthetase F from *E. coli* (P11454); Dhbf, 2,3-dihydroxybenzoate synthetase F from *B. subtilis* (P45745); FenD, fengycin synthetase FenD1 from *B. subtilis* (AJ011849); SrfAA, SrfAB, and SrfAC, surfactin A synthetase A, B, and C, respectively, from *B. subtilis* (X70356); XabB, albicidin synthase B from *X. albilineans* (this study). The A4 to A5 regions (about 100 aa) of adenylation domains of peptide synthetases, which is involved in amino acid recognition and binding, were aligned using the PILEUP program with default parameters.

Figure 8 is a diagrammatic representation showing a restriction map of clones including the *xabA* gene from *X. albilineans*. Sequencing by primer walking commenced at the T3 and T7 primers. The location and direction of transcription of the *xabA* ORF is shown by an arrow. Restriction enzymes are: E, *EcoRI*; P, *PstI*; C, *Clal*; and H, *HindIII*.

Figure 9 is a diagrammatic representation presenting the sequence of the *xabA* gene. The nucleotide sequence is numbered according to the 3-kb sequence in GenBank accession no. AF191324. The closest matches to RBS region and promoter consensus sequences are underlined, as are the region of dyad symmetry and putative factor-independent termination sites. Translation start and stop codons are indicated by boldface type. The (V/I)G(V/I)D and (F/W)(S/C/T)xKE(A/S)xxK motifs conserved in PPTase enzymes are boxed. The insertion site of Tn5 is marked (▼).

Figure 10 is a graphical representation showing albicidin production by wild-type *X. albilineans* strain Xa13 (○), Xa13 pLXABA (●), and complemented Tox⁻ mutant strain LS156 pLXABA (▲). Albicidin concentrations in culture supernatants were quantified based on inhibition zone width in a microbial bioassay (means +/- standard errors from 2 replicates).

Figure 11 is a schematic representation showing a dendrogram (GCG) analysis of PPTases involved in antibiotic and fatty acid biosynthesis in bacteria. Sau, *Salmonella austin*; Sty, *Salmonella typhimurium*; Bbr, *Bacillus brevis*; Xal, *Xanthomonas albilineans*; Eco, *Escherichia coli*; Sfl, *Shigella flexneri*; Bpu, *Bacillus pumilus*; Bsu, *Bacillus subtilis*;
 5 Mtu, *Mycobacterium tuberculosis*; Hin, *Haemophilus influenzae*. The sources of amino acid sequence of PPTases correspond to those in Table 2, and the sequences were aligned using the PILEUP program with default parameters.

Figure 12 is a schematic representation showing the organisation of part of the albicidin biosynthetic gene cluster. The location and direction of three ORFs are indicated
 10 by thick arrows. Vertical lines indicate the position of restriction enzyme sites: E, *EcoRI*; B, *BamHI*; S, *SpeI*; N, *NcoI*. The vertical lines with triangles (\blacktriangle) show the position of insertional mutagenesis sites or Tn5 insertion site, and the resultant mutants are bracketed. The arrows above the physical map indicate the locations of primers used to amplify sequence downstream of the *EcoRI* restriction site by IPCR. The cloned regions for
 15 complementation tests are shown below the map.

Figure 13 is a diagrammatic representation presenting the nucleotide and deduced amino acid sequences of the *xabC* region. The nucleotide sequence is numbered according to the 1515-bp sequence in GenBank accession no. AF239750. The potential RBS and selected restriction sites are underlined. The putative factor-independent termination
 20 signals are underlined and indicated by bold letters. Translation start and stop codons are indicated by bold letters. The conserved motifs in Mtases are boxed. Primers used for PCR (A3F and A3R) and IPCR (IR) are shaded.

Figure 14 is a diagrammatic representation showing the conserved sequence motifs in Mtases involved in antibiotic biosynthesis in bacteria. Identical or similar amino
 25 acids (A = G; D = E; I = L = V) are shown in bold. Numbers indicate amino acid residues from the N terminus of the protein. Xal-XabC, putative albicidin biosynthesis Mtase from *X. albilineans* (this study); Sgl-TcmO and Sgl-TcmN, multifunctional cyclase-dehydrase-3-O-Mtase and tetracenomycin polyketide synthesis 8-O-Mtase of *Streptomyces glaucescens*, respectively (accession number M80674); Smy-MdmC, midecamycin-O-
 30 Mtase of *S. mycarofaciens* (M93958); Mxa-SafC, saframycin O-Mtase of *Myxococcus xanthus* (U24657); Ser-EryG, erythromycin biosynthesis O-Mtase of *Saccharopolyspora*

erythraea (S18533); Spe-DauK, carminomycin 4-*O*-Mtase from *S. peucetius* (L13453); Sal-DmpM, *O*-demethylpuromycin-*O*-Mtase from *S. alboniger* (M74560); Shy-RapM, rapamycin *O*-Mtase of *S. hygroscopicus* (X86780); Sav-AveD, avermectin B 5-*O*-Mtase from *S. avermitilis* (G5921167).

- 5 Figure 15 is a graphical representation showing albicidin production by wild-type *X. albilineans* LS155 (●), Tox⁻ *xabC* insertion mutant LS-JP2 (■), complemented strain LS-JP2 pLXABC containing Lac promoter – full length *xabC* gene (○), and complemented strain LS-JP2 pLXABB1 containing full length *xabB* plus functional N-terminal region of *xabC* (□). Albicidin concentrations in culture supernatants were quantified based on
- 10 inhibition zone width in a microbial bioassay (means +/- standard errors from 2 or 3 replicates).

BRIEF DESCRIPTION OF THE SEQUENCES: SUMMARY TABLE

TABLE A

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 1	Full-length <i>xabB</i> (Accession No. AF239749)	16551 bases
SEQ ID NO: 2	Full-length polypeptide sequence encoded by SEQ ID NO: 1	4801 residues
SEQ ID NO: 3	Full-length coding sequence of <i>xabB</i>	14406 bases
SEQ ID NO: 4	Polypeptide sequence encoded by SEQ ID NO: 3	4801 residues
SEQ ID NO: 5	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl-CoA ligase subdomain I	45 bases
SEQ ID NO: 6	Acyl-CoA ligase subdomain I encoded by SEQ ID NO: 5	15 residues
SEQ ID NO: 7	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl-CoA ligase subdomain II	24 bases
SEQ ID NO: 8	Acyl-CoA ligase subdomain I encoded by SEQ ID NO: 7	8 residues
SEQ ID NO: 9	Sub-sequence of SEQ ID NO: 1 and 3 encoding β -ketoacyl synthase 1 subdomain I	51 bases
SEQ ID NO: 10	β -Ketoacyl synthase 1 subdomain I encoded by SEQ ID NO: 9	17 residues
SEQ ID NO: 11	Sub-sequence of SEQ ID NO: 1 and 3 encoding β -ketoacyl synthase 1 subdomain II	30 bases
SEQ ID NO: 12	β -Ketoacyl synthase 1 subdomain II encoded by SEQ ID NO: 11	10 residues
SEQ ID NO: 13	Sub-sequence of SEQ ID NO: 1 and 3 encoding β -ketoacyl synthase 1 subdomain III	30 bases
SEQ ID NO: 14	β -Ketoacyl synthase 1 subdomain III encoded by SEQ ID NO: 13	10 residues
SEQ ID NO: 15	Sub-sequence of SEQ ID NO: 1 and 3 encoding β -ketoacyl synthase 2 subdomain I	51 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 16	β -Ketoacyl synthase 2 subdomain I encoded by SEQ ID NO: 15	17 residues
SEQ ID NO: 17	Sub-sequence of SEQ ID NO: 1 and 3 encoding β -ketoacyl synthase 2 subdomain II	30 bases
SEQ ID NO: 18	β -Ketoacyl synthase 2 subdomain II encoded by SEQ ID NO: 17	10 residues
SEQ ID NO: 19	Sub-sequence of SEQ ID NO: 1 and 3 encoding β -ketoacyl synthase 2 subdomain III	30 bases
SEQ ID NO: 20	β -Ketoacyl synthase 2 subdomain III encoded by SEQ ID NO: 19	10 residues
SEQ ID NO: 21	Sub-sequence of SEQ ID NO: 1 and 3 encoding β -ketoacyl reductase domain	93 bases
SEQ ID NO: 22	β -Ketoacyl reductase domain encoded by SEQ ID NO: 21	31 residues
SEQ ID NO: 23	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl carrier protein 1 domain	36 bases
SEQ ID NO: 24	Acyl carrier protein 1 domain encoded by SEQ ID NO: 23	12 residues
SEQ ID NO: 25	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl carrier protein 2 domain	36 bases
SEQ ID NO: 26	Acyl carrier protein 2 domain encoded by SEQ ID NO: 25	12 residues
SEQ ID NO: 27	Sub-sequence of SEQ ID NO: 1 and 3 encoding acyl carrier protein 3 domain	36 bases
SEQ ID NO: 28	Acyl carrier protein 3 domain encoded by SEQ ID NO: 27	12 residues
SEQ ID NO: 29	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain I	18 bases
SEQ ID NO: 30	Adenylation domain subdomain I encoded by SEQ ID NO: 29	6 residues
SEQ ID NO: 31	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain II	33 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 32	Adenylation domain subdomain II encoded by SEQ ID NO: 31	11 residues
SEQ ID NO: 33	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain III	48 bases
SEQ ID NO: 34	Adenylation domain subdomain III encoded by SEQ ID NO: 33	16 residues
SEQ ID NO: 35	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain IV	12 bases
SEQ ID NO: 36	Adenylation domain subdomain IV encoded by SEQ ID NO: 35	4 residues
SEQ ID NO: 37	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain V	21 bases
SEQ ID NO: 38	Adenylation domain subdomain V encoded by SEQ ID NO: 37	7 residues
SEQ ID NO: 39	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain VI	45 bases
SEQ ID NO: 40	Adenylation domain subdomain VI encoded by SEQ ID NO: 39	15 residues
SEQ ID NO: 41	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain VII	18 bases
SEQ ID NO: 42	Adenylation domain subdomain VII encoded by SEQ ID NO: 41	6 residues
SEQ ID NO: 43	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain VIII	60 bases
SEQ ID NO: 44	Adenylation domain subdomain VIII encoded by SEQ ID NO: 43	20 residues
SEQ ID NO: 45	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain IX	21 bases
SEQ ID NO: 46	Adenylation domain subdomain IX encoded by SEQ ID NO: 45	7 residues
SEQ ID NO: 47	Sub-sequence of SEQ ID NO: 1 and 3 encoding adenylation domain subdomain X	18 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 48	Adenylation domain subdomain X encoded by SEQ ID NO: 47	6 residues
SEQ ID NO: 49	Sub-sequence of SEQ ID NO: 1 and 3 encoding peptidyl carrier protein 1 domain	33 bases
SEQ ID NO: 50	Peptidyl carrier protein 1 domain encoded by SEQ ID NO: 49	11 residues
SEQ ID NO: 51	Sub-sequence of SEQ ID NO: 1 and 3 encoding peptidyl carrier protein 2 domain	33 bases
SEQ ID NO: 52	Peptidyl carrier protein 2 domain encoded by SEQ ID NO: 51	11 residues
SEQ ID NO: 53	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain I	30 bases
SEQ ID NO: 54	Condensation domain 1 subdomain I encoded by SEQ ID NO: 53	10 residues
SEQ ID NO: 55	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain II	27 bases
SEQ ID NO: 56	Condensation domain 1 subdomain II encoded by SEQ ID NO: 55	9 residues
SEQ ID NO: 57	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain III	30 bases
SEQ ID NO: 58	Condensation domain 1 subdomain III encoded by SEQ ID NO: 57	10 residues
SEQ ID NO: 59	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain IV	21 bases
SEQ ID NO: 60	Condensation domain 1 subdomain IV encoded by SEQ ID NO: 59	7 residues
SEQ ID NO: 61	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain V	36 bases
SEQ ID NO: 62	Condensation domain 1 subdomain V encoded by SEQ ID NO: 61	12 residues
SEQ ID NO: 63	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain VI	21 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 64	Condensation domain 1 subdomain VI encoded by SEQ ID NO: 63	7 residues
SEQ ID NO: 65	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 1 subdomain VII	24 bases
SEQ ID NO: 66	Condensation domain 1 subdomain VII encoded by SEQ ID NO: 65	8 residues
SEQ ID NO: 67	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain I	30 bases
SEQ ID NO: 68	Condensation domain 2 subdomain I encoded by SEQ ID NO: 67	10 residues
SEQ ID NO: 69	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain II	27 bases
SEQ ID NO: 70	Condensation domain 2 subdomain II encoded by SEQ ID NO: 69	9 residues
SEQ ID NO: 71	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain III	30 bases
SEQ ID NO: 72	Condensation domain 2 subdomain III encoded by SEQ ID NO: 71	10 residues
SEQ ID NO: 73	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain IV	21 bases
SEQ ID NO: 74	Condensation domain 2 subdomain IV encoded by SEQ ID NO: 73	7 residues
SEQ ID NO: 75	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain V	33 bases
SEQ ID NO: 76	Condensation domain 2 subdomain V encoded by SEQ ID NO: 75	11 residues
SEQ ID NO: 77	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain VI	21 bases
SEQ ID NO: 78	Condensation domain 2 subdomain VI encoded by SEQ ID NO: 77	7 residues
SEQ ID NO: 79	Sub-sequence of SEQ ID NO: 1 and 3 encoding condensation domain 2 subdomain VII	24 bases

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 80	Condensation domain 2 subdomain VII encoded by SEQ ID NO: 79	8 residues
SEQ ID NO: 81	Polynucleotide comprising <i>xabB</i> promoter	242 bases
SEQ ID NO: 82	Full-length <i>xabA</i> (Accession No. AF191324)	1200 bases
SEQ ID NO: 83	Full-length polypeptide sequence encoded by SEQ ID NO: 82	278 residues
SEQ ID NO: 84	Full-length coding sequence of <i>xabA</i>	837 bases
SEQ ID NO: 85	Polypeptide sequence encoded by SEQ ID NO: 84	278 residues
SEQ ID NO: 86	Sub-sequence of SEQ ID NO: 82 encoding PPTase domain	168 bases
SEQ ID NO: 87	PPTase domain encoded by SEQ ID NO: 86	56 residues
SEQ ID NO: 88	Sub-sequence of SEQ ID NO: 82 encoding a motif (motif I) conserved in PPTases	27 bases
SEQ ID NO: 89	PPTase motif I amino acid sequence encoded by SEQ ID NO: 88	9 residues
SEQ ID NO: 90	Sub-sequence of SEQ ID NO: 82 encoding intervening amino acid sequence linking motifs I and II	117 bases
SEQ ID NO: 91	Intervening amino acid sequence encoded by SEQ ID NO: 90	39 residues
SEQ ID NO: 92	Sub-sequence of SEQ ID NO: 82 encoding a motif (motif II) conserved in PPTases	36 bases
SEQ ID NO: 93	PPTase motif II amino acid sequence encoded by SEQ ID NO: 92	12 residues
SEQ ID NO: 94	Full-length <i>xabC</i> (Accession No. AF239750)	1515 bases
SEQ ID NO: 95	Full-length polypeptide sequence encoded by SEQ ID NO: 94	343 residues
SEQ ID NO: 96	Full-length coding sequence of <i>xabC</i>	1029 bases
SEQ ID NO: 97	Polypeptide sequence encoded by SEQ ID NO: 96	343 residues

SEQUENCE ID NUMBER	SEQUENCE	LENGTH
SEQ ID NO: 98	Sub-sequence of SEQ ID NO: 94 encoding a motif (motif I) conserved in methyltransferases	21 bases
SEQ ID NO: 99	Methyltransferase motif I amino acid sequence encoded by SEQ ID NO: 98	7 residues
SEQ ID NO: 100	Sub-sequence of SEQ ID NO: 94 encoding a motif (motif II) conserved in methyltransferases	24 bases
SEQ ID NO: 101	Methyltransferase motif II amino acid sequence encoded by SEQ ID NO: 100	8 residues
SEQ ID NO: 102	Sub-sequence of SEQ ID NO: 94 encoding a motif (motif III) conserved in methyltransferases	27 bases
SEQ ID NO: 103	Methyltransferase motif III amino acid sequence encoded by SEQ ID NO: 102	9 residues
SEQ ID NO: 104	Polynucleotide encoding said motifs I, II and III	303 bases
SEQ ID NO: 105	Polypeptide encoded by SEQ ID NO: 104	101 residues
SEQ ID NO: 106	Biologically active fragment of SEQ ID NO: 94	831 bases
SEQ ID NO: 107	Biologically active fragment of SEQ ID NO: 95	277 residues

DETAILED DESCRIPTION OF THE INVENTION

1. Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred methods and materials are described. For the purposes of the present invention, the following terms are defined below.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The term "about" is used herein to refer to sequences that vary by as much as 30%, preferably by as much as 20% and more preferably by as much as 10% to the length of a reference sequence.

By "agent" is meant a naturally occurring or synthetically produced molecule which interacts either directly or indirectly with a target member, the level and/or functional activity of which are to be modulated.

"Amplification product" refers to a nucleic acid product generated by nucleic acid amplification techniques.

By "antigen-binding molecule" is meant a molecule that has binding affinity for a target antigen. It will be understood that this term extends to immunoglobulins, immunoglobulin fragments and non-immunoglobulin derived protein frameworks that exhibit antigen-binding activity.

As used herein, the term "binds specifically" and the like refers to antigen-binding molecules that bind the polypeptide or polypeptide fragments of the invention but do not significantly bind to homologous prior art polypeptides.

- 33 -

By "*biologically active fragment*" is meant a fragment of a full-length parent polypeptide which fragment retains the activity of the parent polypeptide. A biologically active fragment will therefore comprise an activity selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity. As used herein, the term "*biologically active fragment*" includes deletion mutants and small peptides, for example of at least 10, preferably at least 20 and more preferably at least 30 contiguous amino acids, which comprise the above activities. Peptides of this type may be obtained through the application of standard recombinant nucleic acid techniques or synthesised using conventional liquid or solid phase synthesis techniques. For example, reference may be made to solution synthesis or solid phase synthesis as described, for example, in Chapter 9 entitled "*Peptide Synthesis*" by Atherton and Shephard which is included in a publication entitled "*Synthetic Vaccines*" edited by Nicholson and published by Blackwell Scientific Publications. Alternatively, peptides can be produced by digestion of a polypeptide of the invention with proteinases such as endoLys-C, endoArg-C, endoGlu-C and staphylococcus V8-protease. The digested fragments can be purified by, for example, high performance liquid chromatographic (HPLC) techniques.

Throughout this specification, unless the context requires otherwise, the words "*comprise*", "*comprises*" and "*comprising*" will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements.

By "*corresponds to*" or "*corresponding to*" is meant a polynucleotide (a) having a nucleotide sequence that is substantially identical or complementary to all or a portion of a reference polynucleotide sequence or (b) encoding an amino acid sequence identical to an amino acid sequence in a peptide or protein. This phrase also includes within its scope a peptide or polypeptide having an amino acid sequence that is substantially identical to a sequence of amino acids in a reference peptide or protein.

By "*derivative*" is meant a polypeptide that has been derived from the basic sequence by modification, for example by conjugation or complexing with other chemical moieties or by post-translational modification techniques as would be understood in the art.

- 34 -

The term "*derivative*" also includes within its scope alterations that have been made to a parent sequence including additions, or deletions that provide for functionally equivalent molecules. Accordingly, the term derivative encompasses molecules that will have an activity selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity.

"*Homology*" refers to the percentage number of amino acids that are identical or constitute conservative substitutions as defined in Table B *infra*. Homology may be determined using sequence comparison programs such as GAP (Deveraux *et al.* 1984, *Nucleic Acids Research* 12, 387-395). In this way, sequences of a similar or substantially different length to those cited herein might be compared by insertion of gaps into the alignment, such gaps being determined, for example, by the comparison algorithm used by GAP.

"*Hybridisation*" is used herein to denote the pairing of complementary nucleotide sequences to produce a DNA-DNA hybrid or a DNA-RNA hybrid. Complementary base sequences are those sequences that are related by the base-pairing rules. In DNA, A pairs with T and C pairs with G. In RNA U pairs with A and C pairs with G. In this regard, the terms "match" and "mismatch" as used herein refer to the hybridisation potential of paired nucleotides in complementary nucleic acid strands. Matched nucleotides hybridise efficiently, such as the classical A-T and G-C base pair mentioned above. Mismatches are other combinations of nucleotides that do not hybridise efficiently.

Reference herein to "*immuno-interactive*" includes reference to any interaction, reaction, or other form of association between molecules and in particular where one of the molecules is, or mimics, a component of the immune system.

By "*immuno-interactive fragment*" is meant a fragment of a parent or reference polypeptide as described herein, which fragment elicits an immune response, including the production of elements that specifically bind to said polypeptide, or variant or derivative thereof. As used herein, the term "*immuno-interactive fragment*" includes deletion mutants and small peptides, for example of at least six, preferably at least 8 and more preferably at

- 35 -

least 20 contiguous amino acids, which comprise antigenic determinants or epitopes. Several such fragments may be joined together.

By "*isolated*" is meant material that is substantially or essentially free from components that normally accompany it in its native state. For example, an "isolated polynucleotide", as used herein, refers to a polynucleotide, which has been purified from the sequences which flank it in a naturally occurring state, *e.g.*, a DNA fragment which has been removed from the sequences which are normally adjacent to the fragment.

By "*modulating*" is meant increasing or decreasing, either directly or indirectly, the level and/or functional activity of a target molecule. For example, an agent may indirectly modulate the said level/activity by interacting with a molecule other than the target molecule. In this regard, indirect modulation of a gene encoding a target polypeptide includes within its scope modulation of the expression of a first nucleic acid molecule, wherein an expression product of the first nucleic acid molecule modulates the expression of a nucleic acid molecule encoding the target polypeptide.

By "*obtained from*" is meant that a sample such as, for example, a nucleic acid extract or polypeptide extract is isolated from, or derived from, a particular source. For example, the extract may be isolated directly from any organism that produces secondary metabolites, preferably from an albicidin-producing microorganism, more preferably from microorganisms of the genus *Xanthomonas*.

The term "*oligonucleotide*" as used herein refers to a polymer composed of a multiplicity of nucleotide units (deoxyribonucleotides or ribonucleotides, or related structural variants or synthetic analogues thereof) linked via phosphodiester bonds (or related structural variants or synthetic analogues thereof). Thus, while the term "oligonucleotide" typically refers to a nucleotide polymer in which the nucleotides and linkages between them are naturally occurring, it will be understood that the term also includes within its scope various analogues including, but not restricted to, peptide nucleic acids (PNAs), phosphoramidates, phosphorothioates, methyl phosphonates, 2-O-methyl ribonucleic acids, and the like. The exact size of the molecule may vary depending on the particular application. An oligonucleotide is typically rather short in length, generally from about 10 to 30 nucleotides, but the term can refer to molecules of any length, although the term "polynucleotide" or "nucleic acid" is typically used for large oligonucleotides.

- 36 -

By "*operably linked*" is meant that transcriptional and translational regulatory nucleic acids are positioned relative to a polypeptide-encoding polynucleotide in such a manner that the polynucleotide is transcribed and the polypeptide is translated.

The term "*polynucleotide*" or "*nucleic acid*" as used herein designates mRNA,
5 RNA, cRNA, cDNA or DNA. The term typically refers to oligonucleotides greater than 30 nucleotides in length.

The terms "*polynucleotide variant*" and "*variant*" refer to polynucleotides displaying substantial sequence identity with a reference polynucleotide sequence or polynucleotides that hybridise with a reference sequence under stringent conditions that are
10 defined hereinafter. These terms also encompass polynucleotides in which one or more nucleotides have been added or deleted, or replaced with different nucleotides. In this regard, it is well understood in the art that certain alterations inclusive of mutations, additions, deletions and substitutions can be made to a reference polynucleotide whereby the altered polynucleotide retains the biological function or activity of the reference
15 polynucleotide. The terms "*polynucleotide variant*" and "*variant*" also include naturally occurring allelic variants.

"*Polypeptide*", "*peptide*" and "*protein*" are used interchangeably herein to refer to a polymer of amino acid residues and to variants and synthetic analogues of the same. Thus, these terms apply to amino acid polymers in which one or more amino acid residues
20 is a synthetic non-naturally occurring amino acid, such as a chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally-occurring amino acid polymers.

The term "*polypeptide variant*" refers to polypeptides in which one or more amino acids have been replaced by different amino acids. It is well understood in the art
25 that some amino acids may be changed to others with broadly similar properties without changing the nature of the activity of the polypeptide (conservative substitutions) as described hereinafter. These terms also encompass polypeptides in which one or more amino acids have been added or deleted, or replaced with different amino acids. Accordingly, polypeptide variants as used herein encompass polypeptides that have an
30 activity selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl

- 37 -

carrier protein activity, condensation activity, PPTase activity and methyltransferase activity.

By "*primer*" is meant an oligonucleotide which, when paired with a strand of DNA, is capable of initiating the synthesis of a primer extension product in the presence of a suitable polymerising agent. The primer is preferably single-stranded for maximum efficiency in amplification but may alternatively be double-stranded. A primer must be sufficiently long to prime the synthesis of extension products in the presence of the polymerisation agent. The length of the primer depends on many factors, including application, temperature to be employed, template reaction conditions, other reagents, and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15 to 35 or more nucleotides, although it may contain fewer nucleotides. Primers can be large polynucleotides, such as from about 200 nucleotides to several kilobases or more. Primers may be selected to be "substantially complementary" to the sequence on the template to which it is designed to hybridise and serve as a site for the initiation of synthesis. By "substantially complementary", it is meant that the primer is sufficiently complementary to hybridise with a target nucleotide sequence. Preferably, the primer contains no mismatches with the template to which it is designed to hybridise but this is not essential. For example, non-complementary nucleotides may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the template. Alternatively, non-complementary nucleotides or a stretch of non-complementary nucleotides can be interspersed into a primer, provided that the primer sequence has sufficient complementarity with the sequence of the template to hybridise therewith and thereby form a template for synthesis of the extension product of the primer.

"*Probe*" refers to a molecule that binds to a specific sequence or sub-sequence or other moiety of another molecule. Unless otherwise indicated, the term "probe" typically refers to a polynucleotide probe that binds to another nucleic acid, often called the "target nucleic acid", through complementary base pairing. Probes may bind target nucleic acids lacking complete sequence complementarity with the probe, depending on the stringency of the hybridisation conditions. Probes can be labelled directly or indirectly.

- 38 -

The term "*recombinant polynucleotide*" as used herein refers to a polynucleotide formed *in vitro* by the manipulation of nucleic acid into a form not normally found in nature. For example, the recombinant polynucleotide may be in the form of an expression vector. Generally, such expression vectors include transcriptional and translational regulatory nucleic acid operably linked to the nucleotide sequence.

By "*recombinant polypeptide*" is meant a polypeptide made using recombinant techniques, *i.e.*, through the expression of a recombinant polynucleotide.

By "*reporter molecule*" as used in the present specification is meant a molecule that, by its chemical nature, provides an analytically identifiable signal that allows the detection of a complex comprising an antigen-binding molecule and its target antigen. The term "*reporter molecule*" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Terms used to describe sequence relationships between two or more polynucleotides or polypeptides include "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity" and "substantial identity". A "*reference sequence*" is at least 12 but frequently 15 to 18 and often at least 25 monomer units, inclusive of nucleotides and amino acid residues, in length. Because two polynucleotides may each comprise (1) a sequence (*i.e.*, only a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity. A "*comparison window*" refers to a conceptual segment of at least 6 contiguous positions, usually about 50 to about 100, more usually about 100 to about 150 in which a sequence is compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. The comparison window may comprise additions or deletions (*i.e.*, gaps) of about 20% or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by computerised implementations of algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release

- 39 -

7.0, Genetics Computer Group, 575 Science Drive Madison, WI, USA) or by inspection and the best alignment (*i.e.*, resulting in the highest percentage homology over the comparison window) generated by any of the various methods selected. Reference also may be made to the BLAST family of programs as for example disclosed by Altschul *et al.*, 1997, *Nucl. Acids Res.* 25:3389. A detailed discussion of sequence analysis can be found in Unit 19.3 of Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley & Sons Inc, 1994-1998, Chapter 15.

The term "*sequence identity*" as used herein refers to the extent that sequences are identical on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a "*percentage of sequence identity*" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) or the identical amino acid residue (*e.g.*, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. For the purposes of the present invention, "*sequence identity*" will be understood to mean the "match percentage", calculated by the DNASIS computer program (Version 2.5 for windows; available from Hitachi Software engineering Co., Ltd., South San Francisco, California, USA) using standard defaults as used in the reference manual accompanying the software.

"*Stringency*" as used herein, refers to the temperature and ionic strength conditions, and presence or absence of certain organic solvents, during hybridisation and washing procedures. The higher the stringency, the higher will be the degree of complementarity between immobilised target nucleotide sequences and the labelled probe polynucleotide sequences that remain hybridised to the target after washing.

"*Stringent conditions*" refers to temperature and ionic conditions under which only nucleotide sequences having a high frequency of complementary bases will hybridise. The stringency required is nucleotide sequence dependent and depends upon the various components present during hybridisation and subsequent washes, and the time allowed for these processes. Generally, in order to maximise the hybridisation rate, non-stringent

- 40 -

hybridisation conditions are selected; about 20 to 25 °C lower than the thermal melting point (T_m). The T_m is the temperature at which 50% of specific target sequence hybridises to a perfectly complementary probe in solution at a defined ionic strength and pH. Generally, in order to require at least about 85% nucleotide complementarity of hybridised sequences, highly stringent washing conditions are selected to be about 5 to 15 °C lower than the T_m . In order to require at least about 70% nucleotide complementarity of hybridised sequences, moderately stringent washing conditions are selected to be about 15 to 30 °C lower than the T_m . Highly permissive (low stringency) washing conditions may be as low as 50 °C below the T_m , allowing a high level of mis-matching between hybridised sequences. Those skilled in the art will recognise that other physical and chemical parameters in the hybridisation and wash stages can also be altered to affect the outcome of a detectable hybridisation signal from a specific level of homology between target and probe sequences. Other examples of stringency conditions are described in section 3.3.

By "*vector*" is meant a nucleic acid molecule, preferably a DNA molecule derived, for example, from a plasmid, bacteriophage, or plant virus, into which a nucleic acid sequence may be inserted or cloned. A vector preferably contains one or more unique restriction sites and may be capable of autonomous replication in a defined host cell including a target cell or tissue or a progenitor cell or tissue thereof, or be integrable with the genome of the defined host such that the cloned sequence is reproducible. Accordingly, the vector may be an autonomously replicating vector, *i.e.*, a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a linear or closed circular plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into a cell, is integrated into the genome of the recipient cell and replicated together with the chromosome(s) into which it has been integrated. A vector system may comprise a single vector or plasmid, two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon. The choice of the vector will typically depend on the compatibility of the vector with the cell into which the vector is to be introduced. The vector may also include a selection marker such as an antibiotic resistance gene that can be used for selection of suitable transformants. Examples of such resistance genes are well known to those of skill in the art.

As used herein, underscoring or italicising the name of a gene shall indicate the gene, in contrast to its protein product, which is indicated by the name of the gene in the absence of any underscoring or italicising. For example, "*xabB*" shall mean the *xabB* gene, whereas "XabB" shall indicate the protein product of the "*xabB*" gene.

5 2. *Isolated polypeptides, biologically active fragments, polypeptide variants and derivatives*

2.1 Polypeptides of the invention

2.1.1 *Albicidin synthetase*

The present inventor has also isolated a gene (*xabB*) encoding a large modular
10 polyketide synthase (PKS) linked to a non-ribosomal peptide synthetase (NRPS) (predicted Mr 525,695). At 4801 amino acids in length, the product of *xabB* (XabB) is the largest reported PKS-NRPS. Comparison of XabB with available protein sequence databases reveals an N-terminal region (from Met-1 to Asp-3235) similar to many microbial modular PKSs, and a C-terminal region (from Pro-3236 to Asp-4801) similar to NRPSs.
15 Recognisable PKS domains commencing at the N-terminus of XabB, are an acyl-CoA ligase (AL), acyl-carrier protein (ACP1), β -ketoacyl synthase (KS1), and β -ketoacyl reductase (KR), followed by two consecutive ACPs and one KS (Figure 1). The motifs characteristic of these domains are aligned with those from other organisms in Figure 3. The AL domain shows 22-30% identity and 50-60% similarity to prokaryotic and
20 eukaryotic aromatic acid-CoA ligases and long-chain fatty acid-CoA ligases, and contains the conserved adenylation core sequence (SGSSG) and the ATPase motif (TGD). The three ACP domains show up to 39.2% identity and 78.6% similarity to acyl carrier proteins, and all contain a 4'-phosphopantetheinyl binding cofactor box GxDS(I/L) (Hopwood and Sherman, 1990), except that A replaces G in ACP1 (Figure 3). The two KS
25 domains show up to 56.1% identity and 80.8% similarity to β -ketoacyl synthases. Both contain motif GPxxxxxxCSxSL around the active site Cys, and two His residues downstream of the active site Cys, in motifs characteristic of these enzymes (Donadio *et al.*, 1991; Hopwood, 1997; Huang *et al.*, 1998). The KR domain shows up to 27.9% identity and 61.8% similarity to β -ketoacyl reductases, and contains the NAD(P)H binding
30 site GGxGxLG (Scrutton *et al.*, 1990).

At the C-terminus of XabB is an apparent peptide synthetase region linked to the PKS module *via* a peptidyl carrier protein (PCP) domain (Figure 1). The peptide synthetase region shows 31-38% identity and 60-63% similarity with members of the peptide synthetase family. It displays the ordered condensation, adenylation, and PCP domains
5 typical of such multienzymes (Marahiel *et al.*, 1997) followed by an extra condensation domain. The conserved sequences, characteristic of the domains commonly found in peptide synthetases, are compared with those from XabB in Table 2.

In more detail, the full-length amino acid sequence of the *X. albilineans* PKS-NRPS, presented in SEQ ID NO: 2, extends 4801 residues and includes the following
10 sequence signature motifs:

(a) acyl-CoA ligase (AL) motif I extending from about residue 226 to about residue 240, and motif II extending from about residue 486 to about residue 493;

(b) β -ketoacyl synthase 1 (KS1) motif I extending from about residue 897 to about residue 913, motif II extending from about residue 1038 to about residue 1047, and
15 motif III extending from about residue 1080 to about residue 1089;

(c) β -ketoacyl synthase 2 (KS2) motif I extending from about residue 2777 to about residue 2793, motif II extending from about residue 2918 to about residue 2927, and
motif III extending from about residue 2955 to about residue 2964;

(d) β -ketoacyl reductase (KR) motif extending from about residue 1812 to about
20 residue 1842;

(e) acyl carrier protein 1 (ACP1) motif extending from about residue 667 to about residue 678;

(f) acyl carrier protein 2 (ACP2) motif extending from about residue 2484 to about residue 2495;

(g) acyl carrier protein 3 (ACP3) motif extending from about residue 2568 to about
25 residue 2579;

(h) adenylation domain (A) motif I extending from about residue 3806 to about residue 3811, motif II extending from about residue 3851 to about residue 3861, motif
30 III extending from about residue 3917 to about residue 3932; motif IV extending from about residue 3967 to about residue 3970, motif V extending from about residue 4063 to about residue 4069, motif VI extending from about residue 4114 to about residue 4128, motif VII extending from about residue 4152 to about residue 4157, motif VIII

extending from about residue 4170 to about residue 4189, motif IX extending from about residue 4239 to about residue 4245, and motif X extending from about residue 4259 to about residue 4264;

(i) peptidyl carrier protein 1 (PCP1) motif extending from about residue 3261 to about residue 3271;

(j) peptidyl carrier protein 2 (PCP2) motif extending from about residue 4306 to about residue 4316;

(k) condensation domain 1 (C1) motif I extending from about residue 3333 to about residue 3342, motif II extending from about residue 3381 to about residue 3389, and motif III extending from about residue 3456 to about residue 3465, motif IV extending from about residue 3495 to about residue 3501, motif V extending from about residue 3606 to about residue 3617, motif VI extending from about residue 3641 to about residue 3647, motif VII extending from about residue 3658 to about residue 3665; and

(l) condensation domain 2 (C2) motif I extending from about residue 4374 to about residue 4383, motif II extending from about residue 4421 to about residue 4429, and motif III extending from about residue 4498 to about residue 4507, motif IV extending from about residue 4538 to about residue 4544, motif V extending from about residue 4649 to about residue 4659, motif VI extending from about residue 4685 to about residue 4691, motif VII extending from about residue 4701 to about residue 4708.

From the above signature motifs, it can be deduced that XabB commences with an AL domain (residues 1-629) followed by an ACP domain (ACP1, residues 630-731). In other PKS systems, an N-terminal AL is involved in activation and incorporation of 3,4-dihydroxycyclohexane carboxylic acid, 3-amino-5-hydroxy benzoic acid (AHBA), or long-chain fatty acid as a starter (Aparicio *et al.*, 1996; Motamedi and Shafiee, 1998; Tang *et al.*, 1998; Duitman *et al.*, 1999). The second module in XabB contains a KS (residues 732-1165), and a KR (residues 1811-1971) upstream of two ACPs (residues 2457-2522, 2544-2613), but lacks any discernable AT domain (Figure 1). The third module contains a KS (residues 2630-3046) followed by a PCP (residues 3221-3307) at the start of the XabB NRPS region.

Four other fused PKS/NRPS systems (Albertini *et al.*, 1995; Gehring *et al.*, 1998; Duitman *et al.*, 1999; Paitan *et al.*, 1999) are known, three of which lack recognisable AT domains (Figure 6). *Yersinia pestis* HMWP1 contains a typical PKS elongation module

- 44 -

(including AT), and an NRPS module with a terminating TE domain. It is the third protein, following an AL (YbtE) and NRPS (HMWP2) in the biosynthetic apparatus for yersiniabactin (Gehring *et al.*, 1998). *B. subtilis* MycA bears the closest resemblance to XabB, showing PKS initiation and elongation modules linked *via* an amino transferase (AMT) domain to the NRPS region. In *B. subtilis* PksK and *M. xanthus* Tal, the NRPS region precedes the PKS region. Separate AT enzymes encoded elsewhere in the genome may operate in *trans* to load the appropriate acyl groups onto the ACPs in the elongation modules of these PKSs. Candidates are a malonyl-CoA transacylase gene (*fenF*) located immediately upstream of *mycA* (Duitman *et al.*, 1999), and an acyltransferase gene located 20 kb upstream of *tal* (Paitan *et al.*, 1999). Accordingly, it is believed that one or more such *trans*-acting AT enzymes may also be involved in connection with the operation of XabB.

From the characteristics of albicidin, and the architecture of the XabB PKS region (Figure 1), the inventor considers that: (i) the AL couples coenzyme A to a shikimate-derived acyl residue in an ATP-dependent reaction, and loads the activated acyl unit onto the 4'-phosphopantetheine prosthetic arm of ACP1; (ii) an acyl group is loaded onto ACP2 or ACP3 by a separate acyltransferase; (iii) the KS1 domain accepts the acyl residue from ACP1 onto a conserved cysteine residue, then transfers it by decarboxylative condensation onto the acyl group tethered to ACP2 or ACP3; (iv) the tethered chain is modified by KR; (v) the assembled polyketide intermediate is translocated *via* KS2 onto the 4-phosphopantetheine prosthetic arm of PCP1, at the start of the NRPS region.

The A domain in the NRPS region of XabB contains ten conserved sequences (A1 to A10, Table 2) identified as AMP, ATP-Mg binding, adenine binding or ATPase sites (Turgay *et al.*, 1992; Marahiel *et al.*, 1997). In other NRPS systems, A domains select and load a particular amino acid, nonproteinogenic amino, hydroxyl or carboxy acid (Marahiel *et al.*, 1997). Substrate specificity is determined at the binding pocket, consisting of a stretch of about 100 amino acid residues between highly conserved motif A4 and A5 (Conti *et al.*, 1997). Sequence alignments for this region reveal some clusters corresponding with the loaded substrate (Stachelhaus *et al.*, 1999). The A domain from XabB falls in a diverse cluster of NRPS modules involved in loading of His, Leu or aromatic amino acids (Phe and Tyr) in other NRPS systems (Figure 7). Based on the architecture of the XabB NRPS region, it can be inferred that the polyketide intermediate

- 45 -

tethered on PCP1 is accepted by C1 and coupled to the amino, hydroxyl, or carboxy acid preloaded by A onto PCP2. The final condensation domain at the C-terminus of XabB is probably involved in peptide-chain termination and cyclisation, as in enniatin, HC-toxin, rapamycin and FK506 systems (Konz and Marahiel, 1999).

5 2.1.2 *Phosphopantetheinyl transferase associated with albicidin biosynthesis*

The present invention also provides a gene (*xabA*) from *X. albilineans* encoding a phosphopantetheinyl transferase (PPTase) associated with XabB function. In this regard, XabB contains five carrier protein (ACP/PCP) domains, to which the growing polyketide or polypeptide chain could be covalently tethered. Each functional ACP or PCP domain
10 must have a specific serine side chain phosphopantetheinylated by a dedicated PPTase (Lambalot *et al.*, 1996). The product of *xabA* (XabA) fulfils this function and is required for post-translational activation of synthetases in the albicidin biosynthetic pathway.

The full-length amino acid sequence of this *X. albilineans* PPTase, presented in SEQ ID NO: 83, extends 278 residues and includes the sequence signature motifs for
15 PPTases which are located as follows: (I) motif I spanning from about residue 159 to about residue 167; and (II) motif II spanning from about residue 207 to about residue 218, of SEQ ID NO: 83. The sequence intervening between the two motifs extends from about residue 168 to about residue 206 of SEQ ID NO: 83. These conserved sequence motifs and the intervening sequence are presented for convenience in SEQ ID NO: 89, 93 and 91,
20 respectively.

The deduced *xabA* gene product has 56-62 % overall similarity to EntD proteins for enterobactin biosynthesis and 39-56 % overall similarity to other enzymes in the phosphopantetheinyl transferase superfamily. Like *entD*, *xabA* includes rarely used codons, which may impose post-transcriptional control on the rate of gene product formation
25 (Coderre & Earhart, 1989). Codon optimisation of *xabA* may, therefore, be useful for enhancing the production of XabA.

2.1.3 Methyltransferase associated with albicidin biosynthesis

The invention also provides a gene (*xabC*) from *X. albilineans* encoding a methyltransferase enzyme, more particularly an *O*-methyltransferase enzyme, which is

- 46 -

required for albicidin production and which when expressed above natural levels leads to increased levels and/or functional activities of albicidin antibiotics. The full-length amino acid sequence of this *X. albilineans* methyltransferase, presented in SEQ ID NO: 95, extends 343 residues and includes methyltransferase consensus sequence motifs which are located as follows: (I) motif I spanning from about residue 173 to about residue 180; (II) motif II spanning from about residue 236 to about residue 243; and (III) motif III spanning from about residue 266 to about residue 274, of SEQ ID NO: 95. These conserved sequence motifs are presented for convenience in SEQ ID NO: 99, 101 and 103, respectively.

10 2.2 Biologically active fragments

The invention also contemplates biological fragments of the above polypeptides of at least 6 and preferably at least 8 amino acids in length, which comprise an activity associated with the domains described above. For example, biologically active fragments may be produced according to any suitable procedure known in the art. For example, a suitable method may include first producing a fragment of a parent polypeptide as described in Section 2.1 and then testing the fragment for the appropriate biological activity. In one embodiment, the fragment is derived from the albicidin PKS-NRPS of the invention and is tested for an activity selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity and condensation activity.

Any assays that detects or preferably measure such activities is contemplated in the practice of the present invention. The biologically active fragment suitably comprises any one or more of the sequence signature motifs described above, or variants thereof. Preferably, the biologically active fragment comprises all said sequence signature motifs, or variants thereof.

In another embodiment, the fragment is derived from the PPTase of the invention and is tested for PPTase activity according to standard assays known to persons of skill in the art. Suitably, the PPTase catalyses the pantetheinylation, more preferably the phosphopantetheinylation, of proteins involved in antibiotic biosynthesis, preferably albicidin biosynthesis. The biologically active fragment preferably comprises the consensus sequence motifs set forth in SEQ ID NO: 89 and 93, or variant thereof and thus,

- 47 -

more preferably comprises the sequence from about residue 159 to about residue 218, of SEQ ID NO: 83.

In yet another embodiment, the fragment is derived from the methyltransferase of the invention and is tested for methyltransferase activity, preferably *O*-methyltransferase activity and more preferably *S*-adenosylmethionine-dependent *O*-methyltransferase activity. Suitably, the methyltransferase catalyses the transfer of one or more methyl groups to an antibiotic precursor, more preferably an albicidin precursor or an intermediate relating to the biosynthesis of albicidins. The biologically active fragment preferably comprises the consensus sequence motifs set forth in SEQ ID NO: 99, 101 and 103, or variant thereof and thus, more preferably comprises the sequence from about residue 173 to about residue 274 of SEQ ID NO: 95 (*i.e.*, SEQ ID NO: 105), or variant of said sequence. In an especially preferred embodiment, the biologically active fragment comprises the sequence from about residue 1 to about residue 277 of SEQ ID NO: 95 (*i.e.*, SEQ ID NO: 107), or variant of said sequence. An exemplary polynucleotide encoding this sequence is cloned in vector pLXABB described *infra*.

Alternatively, biological activity of the fragment is tested by introducing a polynucleotide from which a fragment of a parent polypeptide can be translated into a cell, and detecting one or more of the above activities, which is indicative of said fragment being a biologically active fragment. In one embodiment, such activity can be assayed by introducing into an albicidin deficient *xabB* *X. albilineans* mutant (*e.g.*, strain LS157 described herein) a polynucleotide from which a PKS-NRPS-associated fragment can be produced and assaying for antibiotic activity using a microbial plate assay, as for instance described in Example 1.

In another embodiment, PPTase activity is assayed by introducing into an albicidin deficient *xabA* *X. albilineans* mutant (*e.g.*, strain LS156 described herein) a polynucleotide from which a PPTase-associated fragment can be produced and assaying for antibiotic activity using a microbial plate assay, as for instance described in Example 2.

In yet another embodiment, methyltransferase activity is assayed by introducing into an albicidin deficient *xabC* *X. albilineans* mutant (*e.g.*, strain LS-JP1 described herein) a polynucleotide from which a methyltransferase-associated fragment can be

produced and assaying for antibiotic activity as for example described herein using a microbial plate assay, as for instance described in Example 3.

2.3 Polypeptide variants

The invention also contemplates polypeptide variants of the polypeptides of the invention wherein said variants have an activity selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity, and methyltransferase activity. Suitable methods of producing polypeptide variants include, for example, producing a modified polypeptide whose sequence is distinguished from a parent polypeptide as described in Section 2.1 or a biologically active fragment thereof by the substitution, deletion and/or addition of at least one amino acid. The modified polypeptide is then tested for one or more of said activities, wherein the presence of that activity indicates that the modified polypeptide is a variant of the parent polypeptide.

In another embodiment, a polypeptide variant is produced by introducing into a cell a polynucleotide from which a modified polypeptide can be translated, and detecting one or more of the activities described above that are associated with the cell, which is indicative of the modified polypeptide being a polypeptide variant.

In general, variants will have at least 60%, more suitably at least 70%, preferably at least 80%, and more preferably at least 90% homology to a polypeptide as for example shown in SEQ ID NO: 4, or a biological fragment thereof. It is preferred that variants display at least 60%, more suitably at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90% and still more preferably at least 95% sequence identity with a parent polypeptide as described in Section 2.1 or a biologically active fragment thereof. In this respect, the window of comparison preferably spans about the full length of the polypeptide or of the biologically active fragment. Suitable variants can be obtained from any secondary metabolite-producing organism, and preferably from an albicidin-producing organism.

Alternatively polypeptide variants according to the invention can be identified either rationally, or *via* established methods of mutagenesis (see, for example, Watson, J.

D. *et al.*, "MOLECULAR BIOLOGY OF THE GENE", Fourth Edition, Benjamin/Cummings, Menlo Park, Calif., 1987). Significantly, a random mutagenesis approach requires no *a priori* information about the gene sequence that is to be mutated. This approach has the advantage that it assesses the desirability of a particular mutant based on its function, and thus does not require an understanding of how or why the resultant mutant protein has adopted a particular conformation. Indeed, the random mutation of target gene sequences has been one approach used to obtain mutant proteins having desired characteristics (Leatherbarrow, R. 1986, *J. Prot. Eng.* 1: 7-16; Knowles, J. R., 1987, *Science* 236: 1252-1258; Shaw, W. V., 1987, *Biochem. J.* 246: 1-17; Gerit, J. A. 1987, *Chem. Rev.* 87: 1079-1105). Alternatively, where a particular sequence alteration is desired, methods of site-directed mutagenesis can be employed. Thus, such methods may be used to selectively alter only those amino acids of the protein that are believed to be important (Craik, C. S., 1985, *Science* 228: 291-297; Cronin, *et al.*, 1988, *Biochem.* 27: 4572-4579; Wilks, *et al.*, 1988, *Science* 242: 1541-1544).

Variant peptides or polypeptides, resulting from rational or established methods of mutagenesis or from combinatorial chemistries may comprise conservative amino acid substitutions. Exemplary conservative substitutions in a polypeptide or polypeptide fragment according to the invention may be made according to the following table.

TABLE B

<i>Original Residue</i>	<i>Exemplary Substitutions</i>
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro

- 50 -

<i>Original Residue</i>	<i>Exemplary Substitutions</i>
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile,
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

Substantial changes in function are made by selecting substitutions that are less conservative than those shown in TABLE B. Other replacements would be non-conservative substitutions and relatively fewer of these may be tolerated. Generally, the substitutions which are likely to produce the greatest changes in a polypeptide's properties are those in which (a) a hydrophilic residue (e.g., Ser or Asn) is substituted for, or by, a hydrophobic residue (e.g., Ala, Leu, Ile, Phe or Val); (b) a cysteine or proline is substituted for, or by, any other residue; (c) a residue having an electropositive side chain (e.g., Arg, His or Lys) is substituted for, or by, an electronegative residue (e.g., Glu or Asp) or (d) a residue having a smaller side chain (e.g., Ala, Ser) or no side chain (e.g., Gly) is substituted for, or by, one having a bulky side chain (e.g., Phe or Trp).

2.4 Polypeptide derivatives

A polypeptide can typically tolerate one or more amino acid deletions and insertions in its amino acid sequence without loss or significant loss of a desired activity. Accordingly, the invention also contemplates derivatives of the parent polypeptides of the invention described in Section 2.1 or biologically active fragments thereof or variants of

- 51 -

these, which include amino acid deletions and/or additions, wherein said derivatives comprise one or more activities selected from the group consisting of acyl-CoA ligase activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity associated with antibiotic biosynthesis, and preferably with albicidin biosynthesis.

Preferred derivatives of the invention include PKS-NRPS molecules with altered activities in one or more respects and thus produce polyketides other than the albicidin natural product(s) of the XabB. A PKS-NRPS derived from XabB by such alteration includes a modular PKS-NRPS (or its corresponding encoding gene(s)) that retains the scaffolding of the utilised portion encoded by the naturally occurring gene. Not all domains or modules need be altered. On the constant scaffold, at least one enzymatic activity is mutated, deleted, replaced, or inserted so as to alter the activity of the resulting PKS-NRPS relative to the original or parent PKS-NRPS. Alteration results when these activities are deleted or are replaced by a different version of the activity, or simply mutated in such a way that a polyketide other than the natural product results from these collective activities. This occurs because there has been a resulting alteration of the starter unit and/or elongation unit, stereochemistry, chain length or cyclisation, and/or reductive or dehydration cycle outcome at a corresponding position in the product polyketide. Where a deleted activity is replaced, the origin of the replacement activity may come from a corresponding activity in a different naturally occurring PKS or PKS-NRPS or from a different region of the albicidin PKS-NRPS. Any or all PKS/NRPS genes may be included in the derivative or portions of any of these may be included, but the scaffolding of the albicidin PKS-NRPS protein is preferably retained in whatever derivative is constructed.

Thus, a PKS-NRPS derived from the albicidin PKS-NRPS includes a PKS-NRPS that contains the scaffolding of all or a portion of XabB. The derived PKS-NRPS also contains at least two elongation modules that are functional and preferably at least three elongation modules. The derived PKS-NRPS also contains mutations, deletions, insertions, or replacements of one or more of the activities of the functional domains or modules of XabB so that the nature of the resulting polyketide is altered. Exemplary embodiments include those wherein a KS or ACP domain has been deleted or replaced by a version of the activity from a different PKS/NRPS or from another location within XabB. Also

- 52 -

contemplated are derivatives where at least one non-condensation cycle enzymatic activity (KR, KR, or A) has been deleted or added or wherein any of these activities has been mutated so as to change the structure of the polyketide synthesised by the PKS.

Other derivatives contemplated by the present invention include fusion of the polypeptides, fragments and polypeptide variants of the invention with other polypeptides or proteins. For example, it will be appreciated that said polypeptides, fragments or variants may be incorporated into larger polypeptides, and that such larger polypeptides may also be expected to have one or more of the activities mentioned above. The polypeptides, fragments or variants of the invention may be fused to a further protein, for example, which is not derived from the original host. The further protein may assist in the purification of the fusion protein. For instance, a polyhistidine tag or a maltose binding protein may be used in this respect as described in more detail below. Other possible fusion proteins are those which produce an immunomodulatory response. Particular examples of such proteins include Protein A or glutathione S-transferase (GST).

Other derivatives contemplated by the invention include, but are not limited to, modification to side chains, incorporation of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the polypeptides, fragments and variants of the invention. Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by acylation with acetic anhydride; acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; amidination with methylacetimidate; carbamoylation of amino groups with cyanate; pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with NaBH_4 ; reductive alkylation by reaction with an aldehyde followed by reduction with NaBH_4 ; and trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS). The carboxyl group may be modified by carbodiimide activation via O-acylisourea formation followed by subsequent derivatisation, by way of example, to a corresponding amide. The guanidine group of arginine residues may be modified by formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal. Sulphydryl groups may be modified by methods such as performic acid oxidation to cysteic acid; formation of mercurial derivatives using 4-chloromercuriphenylsulphonic acid, 4-chloromercuribenzoate; 2-chloromercuri-4-

nitrophenol, phenylmercury chloride, and other mercurials; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; carboxymethylation with iodoacetic acid or iodoacetamide; and carbamoylation with cyanate at alkaline pH. Tryptophan residues may be modified, for example, by alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphonyl halides or by oxidation with N-bromosuccinimide. Tyrosine residues may be modified by nitration with tetranitromethane to form a 3-nitrotyrosine derivative. The imidazole ring of a histidine residue may be modified by N-carbethoxylation with diethylpyrocarbonate or by alkylation with iodoacetic acid derivatives.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include but are not limited to, use of 4-amino butyric acid, 6-aminohexanoic acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 4-amino-3-hydroxy-6-methylheptanoic acid, t-butylglycine, norleucine, norvaline, phenylglycine, ornithine, sarcosine, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acids contemplated by the present invention is shown in TABLE C.

TABLE C

<i>Non-conventional amino acid</i>	<i>Non-conventional amino acid</i>
α -aminobutyric acid	L-N-methylalanine
α -amino- α -methylbutyrate	L-N-methylarginine
aminocyclopropane-carboxylate	L-N-methylasparagine
aminoisobutyric acid	L-N-methylaspartic acid
aminonorbornyl-carboxylate	L-N-methylcysteine
cyclohexylalanine	L-N-methylglutamine
cyclopentylalanine	L-N-methylglutamic acid
L-N-methylisoleucine	L-N-methylhistidine
D-alanine	L-N-methylleucine
D-arginine	L-N-methyllysine
D-aspartic acid	L-N-methylmethionine

<i>Non-conventional amino acid</i>	<i>Non-conventional amino acid</i>
D-cysteine	L-N-methylnorleucine
D-glutamate	L-N-methylnorvaline
D-glutamic acid	L-N-methylornithine
D-histidine	L-N-methylphenylalanine
D-isoleucine	L-N-methylproline
D-leucine	L-N-methylserine
D-lysine	L-N-methylthreonine
D-methionine	L-N-methyltryptophan
D-ornithine	L-N-methyltyrosine
D-phenylalanine	L-N-methylvaline
D-proline	L-N-methylethylglycine
D-serine	L-N-methyl-t-butylglycine
D-threonine	L-norleucine
D-tryptophan	L-norvaline
D-tyrosine	α -methyl-aminoisobutyrate
D-valine	α -methyl- γ -aminobutyrate
D- α -methylalanine	α -methylcyclohexylalanine
D- α -methylarginine	α -methylcyclopentylalanine
D- α -methylassparagine	α -methyl- α -naphthylalanine
D- α -methylasspartate	α -methylpenicillamine
D- α -methylcysteine	N-(4-aminobutyl)glycine
D- α -methylglutamine	N-(2-aminoethyl)glycine
D- α -methylhistidine	N-(3-aminopropyl)glycine
D- α -methylisoleucine	N-amino- α -methylbutyrate
D- α -methylleucine	α -naphthylalanine

<i>Non-conventional amino acid</i>	<i>Non-conventional amino acid</i>
D- α -methyllysine	N-benzylglycine
D- α -methylmethionine	N-(2-carbamylethyl)glycine
D- α -methylornithine	N-(carbamylmethyl)glycine
D- α -methylphenylalanine	N-(2-carboxyethyl)glycine
D- α -methylproline	N-(carboxymethyl)glycine
D- α -methylserine	N-cyclobutylglycine
D- α -methylthreonine	N-cycloheptylglycine
D- α -methyltryptophan	N-cyclohexylglycine
D- α -methyltyrosine	N-cyclodecylglycine
L- α -methylleucine	L- α -methyllysine
L- α -methylmethionine	L- α -methylnorleucine
L- α -methylnorvaline	L- α -methylornithine
L- α -methylphenylalanine	L- α -methylproline
L- α -methylserine	L- α -methylthreonine
L- α -methyltryptophan	L- α -methyltyrosine
L- α -methylvaline	L-N-methylhomophenylalanine
N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine
1-carboxy-1-(2,2-diphenyl-ethyl amino)cyclopropane	

Also contemplated is the use of crosslinkers, for example, to stabilise 3D conformations of the polypeptides, fragments or variants of the invention, using homo-bifunctional cross linkers such as bifunctional imido esters having $(CH_2)_n$ spacer groups with $n = 1$ to $n = 6$, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety or

carbodiimide. In addition, peptides can be conformationally constrained, for example, by introduction of double bonds between C_α and C_β atoms of amino acids, by incorporation of C_α and N_α -methylamino acids, and by formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini
5 between two side chains or between a side chain and the N or C terminus of the peptides or analogues. For example, reference may be made to: Marlowe (1993, *Biorganic & Medicinal Chemistry Letters* 3: 437-44) who describes peptide cyclisation on TFA resin using trimethylsilyl (TMSE) ester as an orthogonal protecting group; Pallin and Tam (1995, *J. Chem. Soc. Chem. Comm.* 2021-2022) who describe the cyclisation of
10 unprotected peptides in aqueous solution by oxime formation; Algin *et al* (1994, *Tetrahedron Letters* 35: 9633-9636) who disclose solid-phase synthesis of head-to-tail cyclic peptides *via* lysine side-chain anchoring; Kates *et al* (1993, *Tetrahedron Letters* 34: 1549-1552) who describe the production of head-to-tail cyclic peptides by three-dimensional solid phase strategy; Tumelty *et al* (1994, *J. Chem. Soc. Chem. Comm.* 1067-
15 1068) who describe the synthesis of cyclic peptides from an immobilised activated intermediate, wherein activation of the immobilised peptide is carried out with *N*-protecting group intact and subsequent removal leading to cyclisation; McMurray *et al* (1994, *Peptide Research* 7: 195-206) who disclose head-to-tail cyclisation of peptides attached to insoluble supports by means of the side chains of aspartic and glutamic acid;
20 Hruby *et al* (1994, *Reactive Polymers* 22: 231-241) who teach an alternate method for cyclising peptides *via* solid supports; and Schmidt and Langer (1997, *J. Peptide Res.* 49: 67-73) who disclose a method for synthesising cyclotetrapeptides and cyclopentapeptides. The foregoing methods may be used to produce conformationally constrained polypeptides that comprise one or more activities selected from the group consisting of acyl-CoA ligase
25 activity, β -ketoacyl synthase activity, β -ketoacyl reductase, acyl carrier protein activity, adenylation activity, peptidyl carrier protein activity, condensation activity, PPTase activity and methyltransferase activity associated with the production of polyketides and particularly albicidins or analogues thereof.

The invention also contemplates polypeptides, fragments or variants of the
30 invention that have been modified using ordinary molecular biological techniques so as to improve their resistance to proteolytic degradation or to optimise solubility properties or to render them more suitable as an immunogenic agent.

3. Polynucleotides of the invention

3.1 Polynucleotides encoding polypeptides of the invention

3.1.1 Albicidin synthetase-encoding polynucleotides

The invention further provides a polynucleotide that encodes a PKS-NRPS
5 polypeptide of the invention, or biologically active fragment thereof, or variant or
derivative of these as defined above. In one embodiment, the polynucleotide comprises the
entire sequence of nucleotides set forth in SEQ ID NO: 1. SEQ ID NO: 1 corresponds to a
16511-bp *X. albilineans xabB* cistron. SEQ ID NO: 3, defines the full-length coding
sequence of *xabB* and encodes various sequence signature motifs at the following
10 nucleotide positions:

(a) acyl-CoA ligase (AL) motif I from about nucleotide 676 to about nucleotide 720,
and motif II from about nucleotide 1456 to about nucleotide 1477;

(b) β -ketoacyl synthase 1 (KS1) motif I from about nucleotide 2689 to about
nucleotide 2739, motif II from about nucleotide 3112 to about nucleotide 3141, and
15 motif III from about nucleotide 3238 to about nucleotide 3267;

(c) β -ketoacyl synthase 2 (KS2) motif I from about nucleotide 8329 to about
nucleotide 8379, motif II from about nucleotide 8752 to about nucleotide 8781, and
motif III from about nucleotide 8863 to about nucleotide 8892;

(d) β -ketoacyl reductase (KR) motif from about nucleotide 5434 to about nucleotide
20 5526;

(e) acyl carrier protein 1 (ACP1) motif from about nucleotide 1999 to about
nucleotide 2034;

(f) acyl carrier protein 2 (ACP2) motif from about nucleotide 7450 to about
nucleotide 7485;

(g) acyl carrier protein 3 (ACP3) motif from about nucleotide 7702 to about
25 nucleotide 7735;

(h) adenylation domain (A) motif I from about nucleotide 11416 to about nucleotide
11433, motif II from about nucleotide 11551 to about nucleotide 11583, motif III from
about nucleotide 11749 to about nucleotide 11796; motif IV from about nucleotide
30 11899 to about nucleotide 11910, motif V from about nucleotide 12187 to about
nucleotide 12207, motif VI from about nucleotide 12340 to about nucleotide 12384,

- 58 -

motif VII from about nucleotide 12454 to about nucleotide 12471, motif VIII from about nucleotide 12508 to about nucleotide 12567, motif IX from about nucleotide 12715 to about nucleotide 12735, and motif X from about nucleotide 12715 to about nucleotide 12792;

5 (i) peptidyl carrier protein 1 (PCP1) motif from about nucleotide 9781 to about nucleotide 9813;

(j) peptidyl carrier protein 2 (PCP2) motif from about nucleotide 12915 to about nucleotide 12948;

10 (k) condensation domain 1 (C1) motif I from about nucleotide 9997 to about nucleotide 10026, motif II from about nucleotide 10141 to about nucleotide 10167, and motif III from about nucleotide 10366 to about nucleotide 10395, motif IV from about nucleotide 10483 to about nucleotide 10503, motif V from about nucleotide 10816 to about nucleotide 10851, motif VI from about nucleotide 10921 to about nucleotide 10941, motif VII from about nucleotide 10972 to about nucleotide 10995; and

15 (l) condensation domain 2 (C2) motif I from about nucleotide 13120 to about nucleotide 13149, motif II from about nucleotide 13261 to about nucleotide 13287, and motif III from about nucleotide 13492 to about nucleotide 13521, motif IV from about nucleotide 13612 to about nucleotide 13632, motif V from about nucleotide 13945 to about nucleotide 13977, motif VI from about nucleotide 14053 to about nucleotide 14073, motif VII from about nucleotide 14101 to about nucleotide 14124.

Those of skill in the art will recognise that, due to the degenerate nature of the genetic code, a variety of polynucleotides differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native polynucleotide sequence encoding the PKS-NRPS of *X. albilineans* is shown herein merely to illustrate a preferred embodiment of the invention, and the invention includes polynucleotides of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention.

3.1.2 PPTase-encoding polynucleotides

30 The invention further provides a polynucleotide that encodes a PPTase polypeptide of the invention, or biologically active fragment thereof, or variant or derivative of these as defined above. In one embodiment, the polynucleotide comprises the

- 59 -

entire sequence of nucleotides set forth in SEQ ID NO: 82. SEQ ID NO: 82 corresponds to a 1200-bp *X. albilineans xabA* cistron. This sequence encodes a PPTase catalytic domain from about nucleotide 475 to about nucleotide 654. This domain comprises two conserved PPTase sequence motifs: (I) motif I encoded by a nucleotide sequence from about nucleotide 475 to about nucleotide 501; and (II) motif II encoded by a nucleotide sequence from about nucleotide 619 to about nucleotide 654, of SEQ ID NO: 82. The intervening amino acid sequence, linking motifs I and II, is encoded by a nucleotide sequence from about nucleotide 502 to about nucleotide 618 of SEQ ID NO: 82. The said nucleotide sequences are presented for convenience in SEQ ID NO: 86, 88, 92 and 90, respectively.

Suitably, the polynucleotide comprises the sequence set forth in SEQ ID NO: 84, which defines the full-length coding sequence of *xabA*. Alternatively, the polynucleotide comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 86, which encodes the PPTase catalytic domain.

3.1.3 Methyltransferase-encoding polynucleotides

The invention further provides a polynucleotide that encodes a methyltransferase polypeptide of the invention, or biologically active fragment thereof, or variant or derivative of these as defined above. In one embodiment, the polynucleotide comprises the entire sequence of nucleotides set forth in SEQ ID NO: 94. SEQ ID NO: 94 corresponds to a 1515-bp *X. albilineans xabC* cistron. This sequence encodes three conserved methyltransferase sequence motifs: (I) motif I encoded by a nucleotide sequence from about nucleotide 565 to about nucleotide 585; (II) motif II encoded by a nucleotide sequence from about nucleotide 741 to about nucleotide 774; and (III) motif III encoded by a nucleotide sequence from about nucleotide 841 to about nucleotide 867, or SEQ ID NO: 94. The said nucleotide sequences are presented for convenience in SEQ ID NO: 98, 100 and 102, respectively. Suitably, the polynucleotide comprises the sequence set forth in SEQ ID NO: 96, which defines the full-length coding sequence of *xabC*. Alternatively, the polynucleotide comprises a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 104 or 106, which encode biologically active fragments as described in Section 2.2.

3.2 Polynucleotide variants

In general, polynucleotide variants according to the invention comprise regions that show at least 60%, more suitably at least 70%, preferably at least 80%, and more preferably at least 90% sequence identity over a reference polynucleotide sequence of identical size ("*comparison window*") or when compared to an aligned sequence in which the alignment is performed by a computer homology program known in the art. What constitutes suitable variants may be determined by conventional techniques. For example, a polynucleotide comprising at least one sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102 and 104 can be altered using any suitable method including conventional recombinant techniques and mutagenesis methods such as random mutagenesis (*e.g.*, transposon mutagenesis), oligonucleotide-mediated (or site-directed) mutagenesis, PCR mutagenesis and cassette mutagenesis of an earlier prepared variant or non-variant version of an isolated polynucleotide of the invention.

Alternatively, polynucleotide sequences variants encoding heterologous PKS/NRPS enzymes for producing PKS-NRPS variants of the invention may be obtained from other secondary metabolite- or polyketide-producing organisms. For example, such variants may be prepared according to the following procedure:

(a) creating primers which are optionally degenerate wherein each comprises a portion of a reference polynucleotide encoding a reference polypeptide or fragment of the invention, preferably encoding at least one sequence selected from the group consisting of SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 83, 87, 89, 91, 93, 95, 99, 101, 103, 105 and 107;

(b) obtaining a nucleic acid extract from a secondary metabolite-producing organism, which is preferably a bacterium, more preferably from a species of the family *Pseudomonadaceae*, more preferably from a *Xanthomonas* species; and

(c) using said primers to amplify, *via* nucleic acid amplification techniques, at least one amplification product from said nucleic acid extract, wherein said amplification product corresponds to a polynucleotide variant.

Suitable nucleic acid amplification techniques are well known to the skilled addressee, and include polymerase chain reaction (PCR) as for example described in Ausubel *et al.* (*supra*); strand displacement amplification (SDA) as for example described in U.S. Patent No 5,422,252; rolling circle replication (RCR) as for example described in Liu *et al.*, (1996, *J. Am. Chem. Soc.* 118:1587-1594 and International application WO 92/01813) and Lizardi *et al.*, (International Application WO 97/19193); nucleic acid sequence-based amplification (NASBA) as for example described by Sooknanan *et al.*, (1994, *Biotechniques* 17:1077-1080); and Q- β replicase amplification as for example described by Tyagi *et al.*, (1996, *Proc. Natl. Acad. Sci. USA* 93: 5395-5400).

Typically, polynucleotide variants that are substantially complementary to a reference polynucleotide are identified by blotting techniques that include a step whereby nucleic acids are immobilised on a matrix (preferably a synthetic membrane such as nitrocellulose), followed by a hybridisation step, and a detection step. Southern blotting is used to identify a complementary DNA sequence; northern blotting is used to identify a complementary RNA sequence. Dot blotting and slot blotting can be used to identify complementary DNA/DNA, DNA/RNA or RNA/RNA polynucleotide sequences. Such techniques are well known by those skilled in the art, and have been described in Ausubel *et al.* (1994-1993, *supra*) at pages 2.9.1 through 2.9.20.

According to such methods, Southern blotting involves separating DNA molecules according to size by gel electrophoresis, transferring the size-separated DNA to a synthetic membrane, and hybridising the membrane-bound DNA to a complementary nucleotide sequence labelled radioactively, enzymatically or fluorochromatically. In dot blotting and slot blotting, DNA samples are directly applied to a synthetic membrane prior to hybridisation as above. An alternative blotting step is used when identifying complementary polynucleotides in a cDNA or genomic DNA library, such as through the process of plaque or colony hybridisation. A typical example of this procedure is described in Sambrook *et al.* ("Molecular Cloning. A Laboratory Manual", Cold Spring Harbour Press, 1989) Chapters 8-12.

Typically, the following general procedure can be used to determine hybridisation conditions. Polynucleotides are blotted/transferred to a synthetic membrane, as described above. A reference polynucleotide such as a polynucleotide of the invention is labelled as

- 62 -

described above, and the ability of this labelled polynucleotide to hybridise with an immobilised polynucleotide is analysed. A skilled addressee will recognise that a number of factors influence hybridisation. The specific activity of radioactively labelled polynucleotide sequence should typically be greater than or equal to about 10^8 dpm/mg to provide a detectable signal. A radiolabelled nucleotide sequence of specific activity 10^8 to 10^9 dpm/mg can detect approximately 0.5 pg of DNA. It is well known in the art that sufficient DNA must be immobilised on the membrane to permit detection. It is desirable to have excess immobilised DNA, usually 10 μ g. Adding an inert polymer such as 10% (w/v) dextran sulfate (MW 500,000) or polyethylene glycol 6000 during hybridisation can also increase the sensitivity of hybridisation (see Ausubel *supra* at 2.10.10).

To achieve meaningful results from hybridisation between a polynucleotide immobilised on a membrane and a labelled polynucleotide, a sufficient amount of the labelled polynucleotide must be hybridised to the immobilised polynucleotide following washing. Washing ensures that the labelled polynucleotide is hybridised only to the immobilised polynucleotide with a desired degree of complementarity to the labelled polynucleotide. It will be understood that polynucleotide variants according to the invention will hybridise to a reference polynucleotide under at least low stringency conditions. Reference herein to low stringency conditions include and encompass from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1 M to at least about 2 M salt for hybridisation at 42° C, and at least about 1 M to at least about 2 M salt for washing at 42° C. Low stringency conditions also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M NaHPO₄ (pH 7.2), 7% SDS for hybridisation at 65° C, and (i) 2xSSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM NaHPO₄ (pH 7.2), 5% SDS for washing at room temperature.

Suitably, the polynucleotide variants hybridise to a reference polynucleotide under at least medium stringency conditions. Medium stringency conditions include and encompass from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5 M to at least about 0.9 M salt for hybridisation at 42° C, and at least about 0.1 M to at least about 0.2 M salt for washing at 55° C. Medium stringency conditions also may include 1% Bovine Serum Albumin (BSA), 1 mM EDTA, 0.5 M NaHPO₄ (pH 7.2), 7% SDS for hybridisation at 65° C, and (i) 2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1 mM EDTA, 40 mM NaHPO₄ (pH 7.2), 5% SDS for washing at 60-65° C.

- 63 -

Preferably, the polynucleotide variants hybridise to a reference polynucleotide under high stringency conditions. High stringency conditions include and encompass from at least about 31% v/v to at least about 50% v/v formamide and from about 0.01 M to about 0.15 M salt for hybridisation at 42° C, and about 0.01 M to about 0.02 M salt for washing at 55° C. High stringency conditions also may include 1% BSA, 1 mM EDTA, 0.5 M NaHPO₄ (pH 7.2), 7% SDS for hybridisation at 65° C, and (i) 0.2 x SSC, 0.1% SDS; or (ii) 0.5% BSA, 1mM EDTA, 40 mM NaHPO₄ (pH 7.2), 1% SDS for washing at a temperature in excess of 65° C.

Other stringent conditions are well known in the art. A skilled addressee will recognise that various factors can be manipulated to optimise the specificity of the hybridisation. Optimisation of the stringency of the final washes can serve to ensure a high degree of hybridisation. For detailed examples, see Ausubel *et al.*, *supra* at pages 2.10.1 to 2.10.16 and Sambrook *et al.* (1989, *supra*) at sections 1.101 to 1.104.

While stringent washes are typically carried out at temperatures from about 42° C to 68° C, one skilled in the art will appreciate that other temperatures may be suitable for stringent conditions. Maximum hybridisation rate typically occurs at about 20° C to 25° C below the T_m for formation of a DNA-DNA hybrid. It is well known in the art that the T_m is the melting temperature, or temperature at which two complementary polynucleotide sequences dissociate. Methods for estimating T_m are well known in the art (see Ausubel *et al.*, *supra* at page 2.10.8).

In general, the T_m of a perfectly matched duplex of DNA may be predicted as an approximation by the formula:

$$T_m = 81.5 + 16.6 (\log_{10} M) + 0.41 (\%G+C) - 0.63 (\% \text{ formamide}) - (600/\text{length})$$

wherein: M is the concentration of Na⁺, preferably in the range of 0.01 molar to 0.4 molar; %G+C is the sum of guanosine and cytosine bases as a percentage of the total number of bases, within the range between 30% and 75% G+C; % formamide is the percent formamide concentration by volume; length is the number of base pairs in the DNA duplex.

- 64 -

The T_m of a duplex DNA decreases by approximately 1°C with every increase of 1% in the number of randomly mismatched base pairs. Washing is generally carried out at $T_m - 15^\circ\text{C}$ for high stringency, or $T_m - 30^\circ\text{C}$ for moderate stringency.

In a preferred hybridisation procedure, a membrane (*e.g.*, a nitrocellulose membrane or a nylon membrane) containing immobilised DNA is hybridised overnight at 42°C in a hybridisation buffer (50% deionised formamide, 5xSSC, 5x Denhardt's solution (0.1% ficoll, 0.1% polyvinylpyrrolidone and 0.1% bovine serum albumin), 0.1% SDS and 200 mg/mL denatured salmon sperm DNA) containing labelled probe. The membrane is then subjected to two sequential medium stringency washes (*i.e.*, 2xSSC, 0.1% SDS for 15 min at 45°C , followed by 2xSSC, 0.1% SDS for 15 min at 50°C), followed by two sequential higher stringency washes (*i.e.*, 0.2xSSC, 0.1% SDS for 12 min at 55°C followed by 0.2xSSC and 0.1% SDS solution for 12 min at $65-68^\circ\text{C}$).

Methods for detecting a labelled polynucleotide hybridised to an immobilised polynucleotide are well known to practitioners in the art. Such methods include autoradiography, phosphorimaging, and chemiluminescent, fluorescent and colorimetric detection.

4. Expression vectors

The present invention further provides expression vectors designed for genetic transformation of cells, preferably prokaryotic cells, comprising a polynucleotide, fragment or variant according to the invention operably linked to a regulatory polynucleotide. An expression vector is typically a nucleic acid that can be introduced into a host cell or cell-free transcription and translation system. An expression vector can be maintained permanently or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm.

The various components of an expression vector can vary widely, depending on the intended use of the vector and especially the host cell(s) in which the vector is intended to replicate or drive expression. For example, the regulatory polynucleotide, which is used to control expression of a polynucleotide of the invention, will generally be appropriate for the host cell used for expression. Numerous types of appropriate expression vectors and suitable regulatory sequences are known in the art for a variety of host cells. Typically, the

- 65 -

regulatory polynucleotide includes, but is not limited to, promoter sequences, leader or signal sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and termination sequences, and enhancer or activator sequences. Constitutive or inducible promoters as known in the art are contemplated by the invention.

- 5 The promoters may be either naturally occurring promoters, or hybrid promoters that combine elements of more than one promoter.

In a preferred embodiment, the expression vector is operable in a Gram-negative prokaryotic cell. A variety of prokaryotic expression vectors, which may be used as a basis for constructing the expression vector of the invention. These include but are not limited to
10 a chromosomal vector (e.g., a bacteriophage such as bacteriophage λ), an extrachromosomal vector (e.g., a plasmid or a cosmid expression vector). The expression vector will also typically contain an origin of replication, which allows autonomous replication of the vector, and one or more selectable marker genes that allow phenotypic selection of the transformed cells.

15 The expression vector may also include a fusion partner (typically provided by the expression vector) so that a recombinant polypeptide is expressed as a fusion polypeptide with said fusion partner. The main advantage of fusion partners is that they assist identification and/or purification of said fusion polypeptide. In order to express said fusion polypeptide, it is necessary to ligate a polynucleotide according to the invention into the
20 expression vector so that the translational reading frames of the fusion partner and the polynucleotide coincide. Well known examples of fusion partners include, but are not limited to, glutathione-S-transferase (GST), Fc portion of human IgG, maltose binding protein (MBP) and hexahistidine (HIS₆), which are particularly useful for isolation of the fusion polypeptide by affinity chromatography. For the purposes of fusion polypeptide
25 purification by affinity chromatography, relevant matrices for affinity chromatography are glutathione-, amylose-, and nickel- or cobalt-conjugated resins respectively. Many such matrices are available in "kit" form, such as the QIAexpress™ system (Qiagen) useful with (HIS₆) fusion partners and the Pharmacia GST purification system. In a preferred embodiment, the recombinant polynucleotide is expressed in the commercial vector
30 pFLAG as described more fully hereinafter. Another fusion partner well known in the art is green fluorescent protein (GFP). This fusion partner serves as a fluorescent "tag" which allows the fusion polypeptide of the invention to be identified by fluorescence microscopy

- 66 -

or by flow cytometry. The GFP tag is useful when assessing subcellular localisation of the fusion polypeptide of the invention, or for isolating cells which express the fusion polypeptide of the invention. Flow cytometric methods such as fluorescence activated cell sorting (FACS) are particularly useful in this latter application. Preferably, the fusion partners also have protease cleavage sites, such as for Factor X_a or Thrombin, which allow the relevant protease to partially digest the fusion polypeptide of the invention and thereby liberate the recombinant polypeptide of the invention therefrom. The liberated polypeptide can then be isolated from the fusion partner by subsequent chromatographic separation. Fusion partners according to the invention also include within their scope "epitope tags", which are usually short peptide sequences for which a specific antibody is available. Well known examples of epitope tags for which specific monoclonal antibodies are readily available include c-Myc, influenza virus, haemagglutinin and FLAG tags.

Preferred host cells for purposes of selecting vector components for expression vectors of the present invention include fungal host cells such as yeast and prokaryotic host cells such as *E. coli* and *X. albilineans*, but mammalian cell cultures can also be used. In hosts such as yeasts, plants, or mammalian cells that ordinarily do not produce modular polyketide synthase enzymes, it may be necessary to provide, also typically by recombinant means, suitable holo-ACP synthases to convert the recombinantly produced PKS to functionality.

The expression vector may be used to transform the desired host cell to produce a recombinant host cell for producing *inter alia* a recombinant polypeptide or polyketides, particularly albicidins or analogues thereof, as described hereinafter.

5. *Methods of preparing the polypeptides of the invention*

Polypeptides of the inventions, including the full-length parent polypeptides described in Section 2.1, or their biologically active fragments comprising, for example one or more domains (or fragments of such domains), or variants or derivatives of these, may be prepared by any suitable procedure known to those of skill in the art. For example, the polypeptides may be prepared by a procedure including the steps of: -

(a) preparing a recombinant polynucleotide comprising a nucleotide sequence encoding a polypeptide comprising the sequence set forth in any one of SEQ ID NO: 4 or a fragment thereof comprising at least one sequence selected from the group

- 67 -

consisting of SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 83, 87, 89, 91, 93, 95, 99, 101, 103, 105 and 107, or variant or derivative of these, which nucleotide sequence is operably linked to a regulatory polynucleotide;

- 5 (b) introducing the recombinant polynucleotide into a suitable host cell;
(c) culturing the host cell to express recombinant polypeptide from said recombinant polynucleotide; and
(d) isolating the recombinant polypeptide.

Suitably, said nucleotide sequence comprises at least one sequence selected from
10 the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102 and 104.

The recombinant polynucleotide is preferably in the form of an expression vector, which includes a self-replicating extra-chromosomal vector such as a plasmid, or a vector
15 that integrates into a host genome, as for example described above in Section 4. The step of introducing the recombinant polynucleotide into the host cell may be effected by any suitable means including transfection, and transformation, the choice of which will be dependent on the host cell employed. Such methods are well known to those of skill in the art.

20 Recombinant polypeptides of the invention may be produced by culturing a host cell transformed with an expression vector containing nucleic acid encoding a polypeptide, biologically active fragment, variant or derivative according to the invention. The conditions appropriate for protein expression will vary with the choice of expression vector and the host cell. This is easily ascertained by one skilled in the art through routine
25 experimentation.

Suitable host cells for expression may be prokaryotic or eukaryotic. One preferred host cell for expression of a polypeptide according to the invention is a bacterium. The bacterium used may be *Escherichia coli*. Alternatively, the host cell may be an insect cell such as, for example, *SF9* cells that may be utilised with a baculovirus expression system.

- 68 -

The recombinant protein may be conveniently prepared by a person skilled in the art using standard protocols as for example described in Sambrook, *et al.*, MOLECULAR CLONING. A LABORATORY MANUAL (Cold Spring Harbor Press, 1989), in particular Sections 16 and 17; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (John Wiley & Sons, Inc. 1994-1998), in particular Chapters 10 and 16; and Coligan *et al.*, CURRENT PROTOCOLS IN PROTEIN SCIENCE (John Wiley & Sons, Inc. 1995-1997), in particular Chapters 1, 5 and 6.

Alternatively, the polypeptide, fragments, variants or derivatives of the invention may be synthesised using solution synthesis or solid phase synthesis as described, for example, in Chapter 9 of Atherton and Shephard (*supra*) and in Roberge *et al* (1995, *Science* 269: 202).

6. Antigen-binding molecules

The invention also contemplates antigen-binding molecules that bind specifically to the aforementioned polypeptides, fragments, variants and derivatives. Preferably, an antigen-binding molecule according to the invention is immuno-interactive with any one or more of the amino acid sequences set forth in SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 83, 87, 89, 91, 93, 95, 99, 101, 103, 105 and 107, or variants thereof.

For example, the antigen-binding molecules may comprise whole polyclonal antibodies. Such antibodies may be prepared, for example, by injecting a polypeptide, fragment, variant or derivative of the invention into a production species, which may include mice or rabbits, to obtain polyclonal antisera. Methods of producing polyclonal antibodies are well known to those skilled in the art. Exemplary protocols which may be used are described for example in Coligan *et al.*, CURRENT PROTOCOLS IN IMMUNOLOGY, (John Wiley & Sons, Inc, 1991), and Ausubel *et al.*, (1994-1998, *supra*), in particular Section III of Chapter 11.

In lieu of the polyclonal antisera obtained in the production species, monoclonal antibodies may be produced using the standard method as described, for example, by Köhler and Milstein (1975, *Nature* 256, 495-497), or by more recent modifications thereof as described, for example, in Coligan *et al.*, (1991, *supra*) by immortalising spleen or other

antibody producing cells derived from a production species which has been inoculated with one or more of the polypeptides, fragments, variants or derivatives of the invention.

The invention also contemplates as antigen-binding molecules Fv, Fab, Fab' and F(ab')₂ immunoglobulin fragments. Alternatively, the antigen-binding molecule may be in the form of a synthetic stabilised Fv (scFv) fragment, a disulphide stabilised Fv (dsFv) fragment, a diabody (dAb), a minibody and the like, or may comprise non-immunoglobulin derived, protein frameworks. The antigen-binding molecules of the invention may be used for affinity chromatography in isolating a natural or recombinant polypeptide or biologically active fragment of the invention. For example reference may be made to immunoaffinity chromatographic procedures described in Chapter 9.5 of Coligan *et al.*, (1995-1997, *supra*). The antigen-binding molecules can be used to screen expression libraries for variant polypeptides of the invention as described herein. They can also be used to detect polypeptides, fragments, variants and derivatives of the invention as described hereinafter.

15 7. Identification of modulators

The invention also contemplates a method of screening for an agent that modulates the expression of a gene selected from *xabB*, *xabA*, or *xabC*, or a gene belonging to the same regulatory or biosynthetic pathway as *xabB*, *xabA*, or *xabC*, or a variant of that gene, or that modulates the level and/or functional activity of an expression product of that gene or its variant. The method comprises contacting a preparation comprising said expression product (*e.g.*, polypeptide or transcript), or a biologically active fragment thereof, or variant or derivative of these, or a genetic sequence that modulates the expression of said gene (*e.g.*, the natural promoter relating to said gene, *e.g.*, the *xabB* promoter, comprising the sequence set forth in SEQ ID NO: 81 or complement thereof), with a test agent, and detecting a change in the level and/or functional activity of said polypeptide or biologically active fragment thereof, or variant or derivative, or of a product expressed from said genetic sequence.

Modulators contemplated by the present invention includes agonists and antagonists of gene expression include antisense molecules, ribozymes and co-suppression molecules, as for example described in Section 2. Agonists include molecules which increase promoter activity or interfere with negative mechanisms. Agonists of a gene

include molecules which overcome any negative regulatory mechanism. Antagonists of polypeptides encoded by a gene of interest include antibodies and inhibitor peptide fragments.

Candidate agents encompass numerous chemical classes, though typically they are
5 organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 Dalton. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical
10 carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including, but not limited to: peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogues or combinations thereof.

Small (non-peptide) molecule modulators of a polypeptide according to the
15 invention, or portion, or domain or module thereof are particularly preferred. In this regard, small organic molecules typically have the ability to gain entry into an appropriate cell and affect the expression of a gene (e.g., by interacting with the regulatory region or transcription factors involved in gene expression); or affect the activity of a gene by inhibiting or enhancing the binding of accessory molecules.

20 Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical
25 modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogues. Screening may also be directed to known pharmacologically active compounds and chemical analogues thereof.

Screening for modulatory agents according to the invention can be achieved by any suitable method. For example, the method may include contacting a cell comprising a
30 polynucleotide corresponding to a gene as defined above, with an agent suspected of having said modulatory activity and screening for the modulation of the level and/or

- 71 -

functional activity of a protein encoded by said polynucleotide, or the modulation of the level of an expression product encoded by the polynucleotide, or the modulation of the activity or expression of a downstream cellular target of said protein or said expression product. Detecting such modulation can be achieved utilising techniques including, but not
5 restricted to, ELISA, cell-based ELISA, filter-binding ELISA, inhibition ELISA, Western blots, immunoprecipitation, slot or dot blot assays, immunostaining, RIA, scintillation proximity assays, fluorescent immunoassays using antigen-binding molecule conjugates or antigen conjugates of fluorescent substances such as fluorescein or rhodamine, Ouchterlony double diffusion analysis, immunoassays employing an avidin-biotin or a
10 streptavidin-biotin detection system, and nucleic acid detection assays including reverse transcriptase polymerase chain reaction (RT-PCR).

It will be understood that a polynucleotide from which a target molecule of interest is regulated or expressed may be naturally occurring in the cell which is the subject of testing or it may have been introduced into the host cell for the purpose of testing.
15 Further, the naturally-occurring or introduced sequence may be constitutively expressed – thereby providing a model useful in screening for agents which down-regulate expression of an encoded product of the sequence wherein said down regulation can be at the nucleic acid or expression product level – or may require activation – thereby providing a model
useful in screening for agents that up-regulate expression of an encoded product of the
20 sequence. Further, to the extent that a polynucleotide is introduced into a cell, that polynucleotide may comprise the entire coding sequence which codes for a target polypeptide or it may comprise a portion of that coding sequence (e.g. a domain or module as herein described) or a portion that regulates expression of a product encoded by the polynucleotide (e.g., a promoter). For example, the promoter that is naturally associated
25 with the polynucleotide (ie. the *xabB* promoter) may be introduced into the cell that is the subject of testing. In this regard, where only the promoter is utilised, detecting modulation of the promoter activity can be achieved, for example, by operably linking the promoter to a suitable reporter polynucleotide including, but not restricted to, green fluorescent protein (GFP), luciferase, β -galactosidase and catecholamine acetyl transferase (CAT). Modulation
30 of expression may be determined by measuring the activity associated with the reporter polynucleotide.

- 72 -

In another example, the subject of detection could be a downstream regulatory or biosynthetic target of the target molecule, rather than target molecule itself or the reporter molecule operably linked to a promoter of a gene encoding a product the expression of which is regulated by the target protein.

5 These methods provide a mechanism for performing high throughput screening of putative modulatory agents such as proteinaceous or non-proteinaceous agents comprising synthetic, combinatorial, chemical and natural libraries. These methods will also facilitate the detection of agents which bind either the polynucleotide encoding the target molecule or which modulate the expression of an upstream molecule, which subsequently modulates
10 the expression of the polynucleotide encoding the target molecule. Accordingly, these methods provide a mechanism of detecting agents that either directly or indirectly modulate the expression and/or activity of a gene or expression product according to the invention.

8. *Production of secondary metabolites*

15 The present invention further relates to a process for enhancing the level and/or functional activity of secondary metabolites, preferably albicidins, using one or more agents selected from the polynucleotides, polypeptides, fragments, variants, derivatives, vectors and modulatory agents described above. The process in a preferred embodiment, includes the steps of stably transforming a host cell with an expression vector as broadly
20 described above, comprising at least one nucleic acid sequence encoding a polypeptide of the invention or a biologically active fragment or variant or derivative of these and isolating transformants which produce an enhanced amount of antibiotics, which are preferably of the albicidin class. The vector optionally comprises a signal sequence for secretion recognised by the host cell. Illustrative secretory leaders include the secretory
25 leaders of penicillinase, α -factor, immunoglobulin, T-cell receptors, outer membrane proteins, glucoamylase, fungal amylase and the like. By fusion in proper reading frame, the mature polypeptide may be secreted into the medium. The host cell may be a eukaryote or a prokaryote cell. In one embodiment, the cell naturally produces polyketides, preferably antibiotic polyketides and, in this regard, the cell is preferably *X. albilinears* or other
30 bacteria capable of producing albicidins. Optionally, the construct may include a transcription regulating sequence, which is not subject to repression by substances present

in the growth medium. The above process may be used to prepare antibiotics directly or they may be used to prepare cell free extracts containing increased quantities of antibiotics, preferably of the albicidin class, for *in vitro* preparation of said antibiotics. Suitably, these cell free extracts may be prepared for example using the method disclosed by Dobrogosz, W.J. (1981) Enzymatic activity. In Manual of Methods for General Bacteriology (Gerhardt, P., ed) Washington, DC: American Society for Microbiology, pp. 365-392. In a preferred embodiment, a vector from which a phosphopantetheinyl transferase (PPTase) can be translated is also introduced into the host cell. Expression of PPTase polynucleotides has been shown to be important for the production of polyketides in heterologous expression systems. Preferably, the PPTase is selected from EntD and/or XabA as for example disclosed herein. If desired, a vector from which a methyltransferase, more preferably an *O*-methyltransferase, and even more preferably an *S*-adenosylmethionine *O*-methyltransferase can be translated may also be introduced into the host cell. An exemplary methyltransferase for this purpose is XabC as described herein.

Alternatively, the expression hosts may be used as a source of increased quantities of antibiotics, which can be subsequently purified as for example disclosed by Birch *et al.* in U.S. Patent No. 4,525,354.

The invention also contemplates use of the polynucleotides, polypeptides, fragments, variant and derivatives of the invention in methods of combinatorial biosynthesis of novel antibiotics as for example disclosed by Khosla *et al.* in U.S. Patent No. 5,712,146, Peterson *et al.* in U.S. Patent No. 5,783,431 and Betlach *et al.* in U.S. Patent No. 6,251,636 or in methods of producing antibiotics in hosts that ordinarily do not produce them as for example disclosed by Barr *et al.* in U.S. Patent No. 6,033,883. As discussed in Section 2.4, the invention contemplates albicidin PKS-NRPS derivatives with altered activities in one or more respects for the production of polyketides other than the albicidin natural product(s) of the XabB. In this regard, expression vectors containing nucleotide sequences encoding a variety of such derivatives for the production of different polyketides are transformed into the appropriate host cells to construct a library. In one embodiment, a mixture of such vectors is transformed into selected host cells and the resulting cells plated into individual colonies and selected to identify successful transformants. A variety of strategies is available to obtain a multiplicity of colonies each containing a PKS gene cluster derived from the naturally occurring host gene cluster so

- 74 -

that each colony in the library produces a different PKS and ultimately a different polyketide, as for example disclosed by Betlach *et al.* in U.S. Patent No. 6,251,636. The libraries thus produced can be considered at four levels: (1) a multiplicity of colonies each with a different PKS-NRPS encoding sequence; (2) the proteins produced from the coding sequences; (3) the polyketides produced from the proteins assembled into a functional PKS-NRPS; and (4) antibiotics or compounds with other desired activities derived from the polyketides. Colonies in the library can be induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of polyketides. Polyketides that are secreted into the media or have been otherwise isolated can be screened for binding to desired targets, such as receptors, signalling proteins, and the like. The supernatants *per se* can be used for screening, or partial or complete purification of the polyketides can first be effected. Typically, such screening methods involve detecting the binding of each member of the library to receptor or other target ligand. Binding can be detected either directly or through a competition assay. Means to screen such libraries for binding are well known in the art. Alternatively, individual polyketide members of the library can be tested against a desired target. In this event, screens wherein the biological response of the target is measured can more readily be included. Antibiotic activity can be verified using typical screening assays such as those for albicidin set forth in Example 1.

The invention also extends to the use of the polynucleotides, polypeptides, fragments, variant and derivatives of the invention for the synthesis of antibiotics, preferably antibiotics of the albicidin class.

The polynucleotides of the invention encoding XabB, or a biologically-active fragment or variant thereof, together with a recombinant polynucleotide encoding a PPTase and/or an *O*-methyltransferase which participate or which are capable of participating in the albicidin biosynthetic pathway, provide the means to engineer high level co-expression of the albicidin synthetase, its activating PPTase and modifying methyltransferase to obtain higher yields of albididins.

In order that the invention may be readily understood and put into practical effect, particular preferred embodiments will now be described by way of the following non-limiting examples.

EXAMPLES

EXAMPLE 1

Albicidin multifunctional synthase gene

Materials and Methods

5 Bacterial strains and plasmids

The properties of bacteria and plasmids used in this example are listed in Table 1.

Media, culture conditions and antibiotics

X. albilineans strains were routinely cultured on SP medium (Birch & Patil, 1985b) at 28° C. *Escherichia coli* DH5α and JM109 were used as hosts in cloning
10 experiments and were grown on LB medium at 37° C (Sambrook *et al.*, 1989). Broth cultures were aerated by shaking at 200 r.p.m. on an orbital shaker. Modified YEB medium (Van Larebeke *et al.*, 1977) for patch mating consisted of 10 mg mL⁻¹ peptone, 5 mg mL⁻¹ yeast extract, 5 mg mL⁻¹ NaCl, 5 mg mL⁻¹ sucrose and 0.5 mg mL⁻¹ MgSO₄·7H₂O. The
following antibiotics were added to media as required: 50 µg kanamycin mL⁻¹; 15 µg
15 tetracycline mL⁻¹; 100 µg ampicillin mL⁻¹.

Routine genetic procedures

Bacterial genomic DNA and plasmid DNA isolation, gel electrophoresis, DNA restriction digests, ligation reactions and transformation were performed by routine procedures (Sambrook *et al.*, 1989). DNA fragments were excised from agarose gels and
20 residual agarose was removed with the BRESAclean™ DNA purification kit (GeneWorks, Adelaide).

Construction of a *X. albilineans* partial genomic library

Genomic DNA from *X. albilineans* Xa13 was digested with *Eco*RI and size-fractionated. DNA fragments of 15 to 20 kb were ligated to dephosphorylated *Eco*RI-cleaved pBluescript SK II. The ligated DNA was electroporated into *E. coli* TOP10.
25

- 76 -

Transformants were selected on LB agar medium containing ampicillin, and stored in LB broth with 15% glycerol at -70°C.

PCR amplification

*Bam*HI-digested genomic DNA from *X. albilineans* LS157 was religated at low concentration (0.5 µg/mL) to generate circular DNA molecules as templates for inverse PCR. Three primers, one from the IS terminal region of Tn5 (IR2: 5'-CGGGATCCTCACATGGAAG TCAGATCCTG-3'), and two flanking the unique *Bam*HI restriction site of Tn5 (BL: 5'-GGGGACCTTGCACAGATAGC-3', and BR: 5'-CATTCCTGTAGCGGATGGAGATC-3'), were used to amplify the sequences flanking the Tn5 insertion in the genome of LS157. The amplified fragments (1.4-kb and 6.0-kb) were cloned into pZerO-2, yielding pZIL and pZIR (Figure 1).

PCR was performed in a volume of 50 µl with 200 ng of genomic DNA (or 10 ng of plasmid DNA), 0.4 ng/µL of each of primer, 0.2 mM of each dNTP, 1.8 mM Mg²⁺, and 1 unit of elongase enzyme mix (Life Technologies). A 10-min initial denaturation step at 94° C was followed by 35 thermal cycles of denaturation at 94° C for 1 min, annealing at 55° C for 1 min, and extension at 72° C for 1 min per 1 kb of expected amplification.

Construction of promoter probes and glucuronidase assay

Plasmid pRG960sd contains a promoterless β-glucuronidase gene (*uidA*) downstream of a multiple cloning site (Van den Edde *et al.*, 1992). Sequence upstream of xabB (nucleotide residues 1005 to 1210 or 521 to 1210) was amplified from pLXABB by PCR. Forward primer P1F1 (5'-ACGCGGATCCCAGCAGGGTGTACATACACG-3'), or P1F2 (5'-TCGCGGATCC GCGCGATTGAAGTAGTCC-3') contained a *Bam*HI restriction site (underlined). Reverse primer P1R (5'-TCCCCCGGGCGGCCAGCGTGGTGCTACTAC-3') introduced a *Xma*I restriction site (underlined). PCR fragments were ligated into *Bam*HI/*Xma*I-cut pRG960sd, yielding pRG960p1 and pRG960p2. These constructs were mobilised from *E. coli* DH5α into *X. albilineans* LS155 as described below.

Promoter strength was quantified by fluorometric analysis of glucuronidase activity (Jefferson, 1987; Xiao *et al.*, 1992). The protein content in cell lysates was determined by the dye-binding method (Bradford, 1976) using a Bio-Rad protein assay kit.

Bacterial conjugation

- 5 DNA transfer between *E. coli* donor (JM109 pLAFR3 \pm insert, or DH5 α pRG960sd \pm insert) and *X. albilineans* recipient (LS157 or LS155) was accomplished by triparental transconjugation with helper strain pRK2013. Mid-log-phase cultures of the recipient were spotted onto agar plates containing YEB medium with no antibiotics (20 μ L per spot). After the liquid was absorbed by the agar, 20 μ L of mid-log-phase culture of the
- 10 helper was added to each spot. The liquid was again allowed to absorb, and 20 μ L of mid-log-phase culture of the donor was added to each spot. After incubation of the mating plates overnight at 28° C, transconjugants were selected on SP plates supplemented with ampicillin, and tetracycline or spectinomycin.

Assay and quantification of albicidin production

- 15 Albicidin was quantified by a microbial plate bioassay as described previously (Birch and Patil, 1985b), except that the 10 mL basal layer of LB agar and the 5mL overlayer of 50% LB with 1% agar were supplemented with tetracycline or spectinomycin, and *E. coli* DH5 α pLAFR3 or pRG960sd was used as the indicator strain. This change avoided interference by tetracycline or spectinomycin, which were added to some cultures
- 20 to ensure retention of pLAFR3 or pRG960sd derivatives in *X. albilineans*. Inhibition zone widths in the bioassay were converted to albicidin concentrations by interpolation on a dose-response plot produced under the same assay conditions. The plot fits the formula: $\text{Log [Alb]} = 0.3 W - 0.92$, where [Alb] is units of albicidin per 20 μ L sample assayed, and W is the width in millimetres of the zone of growth inhibition surrounding each well.

25 *Results*

Cloning and sequencing of *xabB* gene required for albicidin production

Xanthomonas albilineans Tox⁻ mutant LS157 contains a single Tn5 insertion, in a 4.1 kb *Cla*I restriction fragment or a 16.5 kb *Eco*RI restriction fragment (Figure 1).

- 78 -

Selection for kanamycin resistance, following shotgun cloning of *Cla*I restriction fragments of LS157 DNA into pBluescript II SK, yielded clone pBC157. Sequences flanking the Tn5 insertion in LS157 DNA were amplified by inverse PCR, and cloned into pZER0-2, producing pZIL and pZIR. Plasmid pLXABB was screened from a *X. albilineans* *Xa*13 *Eco*RI genomic library with probes described in Figure 1B. Subclones pSEBL and pSEBR were derived from pLXABB (Figure 1C, Table 1).

The double-strand sequence of the 16,511 bp *Eco*RI genomic fragment in pLXABB was obtained by a primer-walking approach, using subclones pBC157, pZIL, pZIR, pSEBL, and pSEBR. The Tn5 insertion in the genome of LS157 is accompanied by 9-bp perfect repeat sequence (GTCCTGAAG), commencing at 2490 bp in GenBank accession no. AF239749.

The only ORF longer than 900 bp within the 16.5-kb fragment is disrupted by the Tn5 insertion. This ORF (designated *xabB*) encodes a protein of 4081 aa (Mr 525,695). It commences at 1230 bp in GenBank accession no. AF239749 with a TTG codon, 6 bp downstream from a ribosome binding sequence (RBS) GAGG, which may impose post-transcriptional control on the rate of gene product formation (McCarthy and Gualerzi, 1990). There is an alternative start codon (ATG) a further 15 bp downstream. Of the codons in this ORF, 8.5% are rarely used in *E. coli*. The closest match (TTGAGC-14x-TATAAC) to the consensus -35 (TTGACA) and -10 (TATAAT) sequences for *E. coli* σ^{70} promoters occurs 117 bp upstream of the translation initiation codon (Figure 2).

Downstream by 35 bp from the TAG stop codon of *xabB* is a probable RBS (GAGG), separated by 6 bp from the ATG start codon of another ORF (designated *xabC*) in the same orientation as *xabB*. Overlapping the *xabB* promoter region is another probable promoter for a divergent transcript including a putative RBS (TGGAGG) and start codon for a gene designated *xatA*, separated by 233 bp from *xabB* (Figure 1, 2).

Complementation of *xabB* gene in LS157

Mobilisation of pLAFR3, pLXABB1 or pLXABB2 by bacterial conjugation into *Tox*⁻ mutant LS157 occurred at a frequency of 1.5×10^{-2} transconjugants/recipient cells. Albicidin production was undetectable in *Tox*⁻ mutant LS157 and LS157 (pLAFR3)

controls, but introduction of the *xabB* gene on pLXABB1 or pLXABB2 restored albicidin production to the level of the wild-type parental strain LS155 (Figure 4).

Functional analysis of *xabB* promoter region

- GUS activity was undetectable in LS155 and LS155 (pRG960sd) controls.
- 5 Plasmid pRG960p1 or pRG960p2, with 206 bp or 690 bp from the *xabB* promoter region upstream of GUS, both conferred GUS activity with no difference in expression level or pattern in *X. albilineans* LS155 (Figure 5).

Discussion

- Albicidin was partially characterised as a low-molecular-weight compound that
- 10 contains 38 carbon atoms with 3-4 aromatic rings (Birch and Patil, 1985a). The compound is not degraded by peptidases (Birch and Patil, 1985a), but it is cleaved by the AlbD esterase (Zhang and Birch, 1997). Based on the deduced functionality of the synthase describe herein, albicidin is likely to be a complex polyketide, condensed with amino acid(s), or nonproteinogenic amino, hydroxyl and carboxyl acid(s) by C-N, amide or ester
- 15 bond formation.

- The characterisation of XabB as a multi-modular hybrid enzyme provides new insights into the mechanism of albicidin biosynthesis and possible approaches to engineer the overproduction of albicidins. For example, the complementation experiments (Figure 4) indicate that increased copy number of *xabB* stimulates early production of albicidin,
- 20 but other factors become limiting during idiophase. It may be possible to increase expression of the albicidin synthase by modifications to the promoter and TTG start codon, or to improve albicidin yields by supplying candidate substrates (such as shikimate-derived units). The unusual enzyme organisation also contributes to the emerging understanding of how microbes generate structural diversity of antibiotics, and can facilitate combinatorial
- 25 engineering of antibiotics of mixed peptide/polyketide origin.

EXAMPLE 2**Albicidin Antibiotic and Phytotoxin Biosynthesis in *Xanthomonas albilineans* Requires a Phosphopantetheinyl Transferase Gene*****Materials and Methods*****5 Bacterial strains and plasmids**

The properties of bacteria and plasmids used in this Example are listed in Table 3.

Media, culture conditions and antibiotics

X. albilineans strains were routinely cultured on SP medium (Birch & Patil, 1985b) at 28° C. *Escherichia coli* DH5 α and JM109 were used as hosts in cloning
10 experiments and were grown on LB medium at 37° C (Sambrook *et al.*, 1989). Broth cultures were aerated by shaking at 200 r.p.m. on an orbital shaker. Modified YEB medium (Van Larebeke *et al.*, 1977) for patch mating consisted of 10 mg mL⁻¹ peptone, 5 mg mL⁻¹ yeast extract, 5 mg mL⁻¹ NaCl, 5 mg mL⁻¹ sucrose and 0.5 mg mL⁻¹ MgSO₄·7H₂O. The following antibiotics were added to media as required: 50 μ g kanamycin mL⁻¹; 15 μ g
15 tetracycline mL⁻¹, 100 μ g ampicillin mL⁻¹.

Assay of albicidin production

Albicidin was quantified by a microbial plate bioassay as described previously (Birch and Patil, 1985b), except that the 10 mL basal layer of LB agar and the 5 mL overlayer of 50% LB with 1% agar were supplemented with tetracycline, and *E. coli*
20 DH5 α [pLAFR3] was used as the indicator strain. This change avoided interference by tetracycline, which was added to some cultures to ensure retention of pLAFR3 derivatives in *X. albilineans*.

Routine genetic procedures

Bacterial genomic DNA and plasmid DNA isolation, gel electrophoresis, DNA
25 restriction digests, ligation reactions and transformation were performed by routine procedures (Sambrook *et al.*, 1989). DNA fragments were excised from agarose gels and

residual agarose was removed with the BRESAclean™ DNA purification kit (GeneWorks, Adelaide).

DNA sequencing and analysis

Sequencing reactions were performed by dideoxynucleotide chain termination (Sanger *et al.*, 1977) using the BigDye™ Terminator Cycle Sequencing Kit and 373A DNA sequencer (PE Applied Biosystems) through the Australian Genome Research Facility. Oligonucleotide primers were purchased from GeneWorks (Adelaide). University of Wisconsin Genetics Computer Group (UWGCG) programs BLASTP, FASTA, PILEUP, and BESTFIT were used through WebANGIS version 2.0 for DNA and protein sequence analyses of the GenBank, EMBL, PIR and SWISSPROT databases using standard defaults.

Cloning of Tn5 flanking sequences

EcoRI-digested genomic DNA from *X. albilineans* Tox⁻ mutant LS156 was ligated into pBluescript II SK and electroporated into *E. coli* DH5α. Transformants were selected on LB medium containing kanamycin and ampicillin, yielding clone pBEA1, from which subclones pCEA1 and pPEA1 were obtained (Figure 1).

Amplification of sequences from wild-type LS155 by PCR

Sequences flanking the Tn5 insertion in LS156 were used to design primers (A1F: 5'-TTTGGGTTGGATCGGGTAG-3' and A1R: 5'-CCTTCTCGTCCTTG CTCTTC-3') for PCR-amplification of the corresponding wild type *X. albilineans* LS155 chromosomal DNA. PCR was performed in a volume of 50 µL with 200 ng of genomic DNA, 0.4 ng µL⁻¹ of each of primer, 0.2 mM of each of dNTP, 1.8 mM Mg²⁺, and 1 unit of elongase enzyme mix (Life Technologies). A 4-min initial denaturation step at 94° C was followed by 35 thermal cycles of denaturation at 94° C for 1 min, annealing at 55° C for 1 min, and extension at 72° C for 2 min. The amplified DNA fragment was cloned into pGEM-T to give pGTA1 (Figure 1).

Construction of expression vectors

The coding region of the *xabA* gene was amplified from pGTA1 by PCR. Primer A1F1 (5'-GGAATTCCATGCCCAATGCCGTACCG-3') contained an *EcoRI* restriction

- 82 -

site (underlined) for insertion of the amplified gene into the correct reading frame of *lacZ* in pLAFR3. Primer A1R1 (5'-CGGGATCCCGTGCTCACCAGGCGTAGTGG-3') introduced a *Bam*HI restriction site (underlined), 5 bases downstream from the stop codon of the amplified gene. The amplified DNA fragment was digested with *Eco*RI and *Bam*HI, and ligated with *Eco*RI/*Bam*HI-digested pLAFR3 to result in pLXABA.

Similarly, the coding region of the *entD* gene was PCR-amplified from *E. coli* DH5 α by colony PCR using primers EntDF (5'-TCCCGGAATTCCATGGTCGATATGAAAACACTACGC-3') and EntDR (5'-GCCCAAGCTTCTAATCGTGTGGCACAGCGTTATG-3'), then ligated into pLAFR3 to produce pLENTD. The inserts in pLXABA and pLENTD were sequenced to confirm the expected clones.

Bacterial triparental mating

DNA transfer between *E. coli* donor (JM109 pLAFR3 \pm insert) and *X. albilineans* recipient (LS155 or LS156) was accomplished by triparental transconjugation with helper strain pRK2013. The mid-log-phase cultures of the recipient were spotted onto agar plates containing YEB medium with no antibiotics (20 μ L per spot). After the liquid was absorbed by the agar, 20 μ L of mid-log-phase culture of the helper was added to each spot. The liquid was again allowed to absorb, and 20 μ L of mid-log-phase culture of the donor was added to each spot. After incubation of the mating plates overnight at 28° C, transconjugants were selected on SP plates supplemented with tetracycline and ampicillin.

Results

Cloning and sequencing of the *xabA* gene required for albicidin production

Xanthomonas albilineans Tox⁻ mutant LS156 contains a single Tn5 insertion, in a 3.0-kb *Eco*RI restriction fragment (Wall & Birch, 1997). Selection for Tn5-encoded kanamycin resistance, following shotgun cloning of *Eco*RI restriction fragments of LS156 DNA into pBluescript II SK, yielded pBEA1 (Figure 8).

Both strands of the insert in pBEA1 excluding the Tn5 insertion were sequenced by primer-walking from T3 and T7 vector sequences in pBEA1 and subclones pCEA1 and

- 83 -

pPEA1. The corresponding genomic region was amplified from wild-type *X. albilineans* LS155 by PCR, and cloned into pGEM-T to give pGTA1. Sequencing of pGTA1 revealed that a 9-bp imperfect repeat sequence (TTGGCCACG) in the genome of LS156 accompanied the Tn5 insertion (following base number 1869 in Figure 9). The double-strand nucleotide sequence of the 2989 bp wild type *EcoRI* fragment is deposited in GenBank under accession no. AF191324.

Reading frame analysis of the 3 kb *EcoRI* fragment revealed that only one ORF (designated xabA) is disrupted by the Tn5 insertion. This ORF encodes a protein of 278 aa (Mr 29 277), with 6.12% codons rarely used in *E. coli*. There were no close matches to *E. coli* -10 (TATAAT) and -35 (TTGACA) consensus promoter sequences, and no appropriately spaced RBS sequence (such as AGGA or GAGG) in the region upstream of the putative start codon ATG (Figure 9). A region of GC-rich dyad symmetry with a free energy of -10.2 kcal/mol was found, followed by two TCTC boxes that closely resemble the TCTG consensus sequence characteristic of many factor-independent termination sites (Brendel & Trifonov, 1984; Platt, 1986) downstream of the TGA termination codon of xabA.

Comparison of XabA with other bacterial PPTases

A search for proteins with homology to the deduced xabA product, using the FASTA and BLASTP and SWISSPROT programs, indicated regions of similarity to EntD from *Escherichia coli* (170 aa overlap, 35.9 % identity, 56.5 % similarity), *Shigella flexneri* (180 aa overlap, 35.0 % identity, 55.6 % similarity), *Salmonella typhimurium* (184 aa overlap, 35.9 % identity, 62.0 % similarity), and *Salmonella austin* (172 aa overlap, 36.1 % identity, 61.1 % similarity). XabA contains (V/I)G(V/I)D and (F/W)(S/C/T)xKE(S/A)xxK domains characteristic of the phosphopantetheinyl transferase (PPTase) superfamily, and shares 17-36 % overall identity, 39-62 % overall similarity, with other bacterial PPTases (Table 4).

Enhanced expression of xabA by complementation in LS156 results in increased production of albicidins

Mobilisation of pLAFR3 or pLXABA (pLAFR3::xabA) by triparental matings into Tox⁻ mutant LS156 occurred at a frequency of 1.5×10^{-2} transconjugants/recipient

cells. Albicidin production was undetectable in *Tox*⁻ mutant LS156 and LS156 (pLAFR3) controls, but introduction of the *xabA* gene on pLXABA enhanced albidin production restored albidin production (Figure 10). In LS156 (pLXABA), as in LS155, albidin was first detectable in late-log-phase cultures ($OD_{550} = 0.7$) and was maximal in stationary phase. Albicidin production was not responsive to IPTG or glucose, and the *lac* promoter driving *xabA* in pLXABA is considered to express constitutively in *X. albilineans*. The *E. coli entD* gene, expressed from the *lac* promoter in pLENTD, also complemented the *xabA::Tn5* mutation, restoring albidin production in LS156.

Discussion

10 A gene required for albidin production in *X. albilineans* was isolated using a Tn5 mutagenesis and shotgun cloning approach. The ORF interrupted by Tn5 in *Tox*⁻ mutant LS156 is designated *xabA*. This ORF was isolated from *Tox*⁺ parent strain LS155, and shown to enhance albidin production early in the production phase in LS156 when expressed from the *lac* promoter in pLAFR3. Tn5 insertions typically cause polar mutations affecting all downstream cistrons in an operon (De Bruijn and Lupski, 1984).
15 Complementation of the mutation in LS156 by the isolated *xabA* ORF indicates the absence of any downstream cistron involved in albidin production. There is no consensus RBS sequence close to the alternative start codons for this ORF in the *X. albilineans* genome. Translation may be initiated without an evident ribosome binding sequence
20 complementary to the 3' end of the 16S rRNA, as observed for some streptomycete genes involved in secondary metabolism (Strohl, 1992), and for some chloroplast genes (Kozak, 1999).

PPTases play an essential role in priming polyketide, fatty acid, non-ribosomal peptide and siderophore biosynthesis (Gehring *et al.*, 1997a; Lambalot *et al.*, 1996; Marahiel *et al.*, 1997; Walsh *et al.*, 1997). All polyketide synthase, fatty acid synthetases, and non-ribosomal peptide synthetases require post-translational modification to become catalytically active (Walsh *et al.*, 1997). The inactive apo-proteins are converted to their active holo-forms by transfer of the 4'-phosphopantetheinyl (P-pant) moiety of coenzyme A to the sidechain hydroxyl of a serine residue in a conserved carrier domain (Lambalot *et al.*, 1996; Walsh *et al.*, 1997). The P-pant moiety serves to covalently tether the growing
30

product, which is assembled by sequential action of multiple catalytic domains in these complex synthetases (Walsh *et al.*, 1997).

A family of more than twenty PPTases is recognised by a common (V/I)G(V/I)D_{x40-45}...(F/W)(S/C/T)_xKE(A/S)_{xx}K signature sequence, but overall sequence homologies are low (Gehring *et al.*, 1997; Lambalot *et al.*, 1996; Nakano *et al.*, 1992; Quadri *et al.*, 1998a). In *E. coli*, there are two PPTases with distinct specificities: ACPS is active on acyl carrier protein (ACP) domains in fatty acid and polyketide synthase; EntD is active on peptidyl carrier protein (PCP) and aryl carrier protein (ArCP) domains in peptide synthetases (Lambalot *et al.*, 1996; Walsh *et al.*, 1997). Thus, PPTases may be partner-protein specific. However, Sfp from *B. subtilis* appears to be non-specific, efficiently activating both fatty acid, polyketide synthase and peptide synthetases (Kealey *et al.*, 1998; Mofid *et al.*, 1999; Quadri *et al.*, 1998a). XabA includes the PPTase VGID and FS_xKES_{xx}K motifs. Although it has highest overall similarity to the peptide-selective EntD proteins, the sequence groupings are not sufficiently compelling to predict the specificity of XabA for polyketide synthase or peptide synthetases (Table 4, Figure 11).

Complementation studies have revealed substantial functional interchangeability of PPTases in different bacteria. For example, the *B. subtilis* *sfp* gene involved in surfactin biosynthesis complements mutants in *E. coli* *entD* (enterobactin biosynthesis) and *B. brevis* *gsp* (gramicidin biosynthesis) (Borchert *et al.*, 1994; Grossman *et al.*, 1993). In vitro, ACPS from *E. coli* activates apoproteins from *Lactobacillus*, *Rhizobium* and *Streptomyces* (Lambalot *et al.*, 1996). Because XabA shows highest similarity to EntD, we amplified the *entD*-coding region from *E. coli*, and arranged it for expression from the *lac* promoter in broad host-range vector pLAFR3. This construct (pLENTD) restored albicidin production in *X. albilineans* *xabA*::Tn5 mutant LS156. EntD is a peptide-selective PPTase that converts inactive apo-EntF and apo-EntB to active holo-enzymes involved in biosynthesis of enterobactin in *E. coli* (Gehring *et al.*, 1997a). Functional complementation of the *xabA*::Tn5 mutation by *entD* indicates that XabA is a PPTase required for post-translational activation of synthetases involved in albicidin production in *X. albilineans*. The specificity of EntD for activation of peptide synthetases in *E. coli* indicates that albicidin biosynthesis probably involves an XabA-activated peptide synthetase.

Some PPTase genes involved in non-ribosomally synthesised peptide biogenesis are located near the genes encoding their targets (Quadri *et al.*, 1998b). For example, *B. brevis gsp*, *B. subtilis sfp*, and *E. coli entD* genes all lie within 4 kb of operons encoding the target peptide synthetases (Borchert *et al.*, 1994; Coderre & Earhart, 1989; Nakano *et al.*, 1992). However, *M. tuberculosis pptT* is not located near the *mbt* gene cluster encoding the target peptide synthetases involved in mycobactin biosynthesis (Quadri *et al.*, 1998b). No gene encoding a PPTase has been identified in any of the antibiotic and phytotoxin biosynthetic gene clusters characterised from *Streptomyces* spp. (Gehring *et al.*, 1997b) and *Pseudomonas* spp. (Bender *et al.*, 1999). No evident target gene was found within 1282 bp upstream or 870 bp downstream of *xabA*. Three cosmids spanning about 100 kb in two regions of the genome complemented 56 of 58 tested *Tox⁻* mutants of *X. albilineans*, but not LS156 (Rott *et al.*, 1996). These results indicate that *xabA* is not clustered with the genes encoding the antibiotic synthetases that it activates.

Expression of *xabA* (or an alternative PPTase such as *entD*) is essential for albicidin biosynthesis. The phosphopantetheinyl transferase gene described herein provides new insight into antibiotic biosynthesis in the *Pseudomonadaceae*, and new opportunities to understand and apply albicidins as potent inhibitors of prokaryote DNA replication. This gene, together with the *xabB* provide the means to engineer high level co-expression of the albicidin synthetase and its activating PPTase to obtain higher yields of albicidins, and ultimately to manipulate the elements of this biosynthetic machinery, by mutagenesis or otherwise, to produce desired structural variants of this novel antibiotic class. They may also indicate a new approach to disease resistance, by engineering plants to interfere with the biosynthesis of albicidin toxins, which are key pathogenesis factors for the systemic development of leaf scald disease.

EXAMPLE 3

A methyltransferase gene is involved in albicidin biosynthesis in *Xanthomonas albilineans*

Material and Methods

Bacterial strains and plasmids

The properties of bacteria and plasmids used in this example are listed in Table 5.

Media, culture conditions and antibiotics

X. albilineans strains were routinely cultured on sucrose peptone (SP) medium at 28° C (Birch and Patil, 1985b). *Escherichia coli* strains were used as hosts in cloning experiments and were grown on LB medium at 37° C (Sambrook *et al.*, 1989). Broth
5 cultures were aerated by shaking at 200 rpm on an orbital shaker. Modified YEB medium (Van Larebere *et al.*, 1977) was used for patch mating. The following antibiotics were added to media as required: kanamycin, 50 µg/mL; tetracycline, 15 µg/mL; ampicillin, 100 µg/mL.

Assay of albicidin production

10 Albicidin was quantified by a microbial plate bioassay as described previously (Birch and Patil, 1985b), except that the 10 mL basal layer of LB agar and the 5 mL overlayer of 50% LB with 1% agar were supplemented with tetracycline, and *E. coli* DH5α [pLAFR3] was used as the indicator strain. This change avoided interference by tetracycline, which was added to some cultures to ensure retention of pLAFR3 derivatives
15 in *X. albilineans*.

Routine genetic procedures

Bacterial genomic DNA and plasmid DNA isolation, gel electrophoresis, DNA restriction digests, ligation reactions and transformation were performed by routine procedures (Sambrook *et al.*, 1989). DNA fragments were excised from agarose gels and
20 residual agarose was removed with the BRESAclean™ DNA purification kit (GeneWorks, Adelaide).

DNA sequencing and analysis

Sequencing reactions were performed by dideoxynucleotide chain termination (Sanger *et al.*, 1977) using the BigDye™ Terminator Cycle Sequencing Kit and 373A
25 DNA sequencer (PE Applied Biosystems) through the Australian Genome Research Facility. Oligonucleotide primers were purchased from GeneWorks (Adelaide). University of Wisconsin Genetics Computer Group (UWGCG) programs BLASTP, FASTA, PILEUP, and BESTFIT were used through WebANGIS version 2.0 for DNA and protein sequence analyses of the GenBank, EMBL, PIR and SWISSPROT databases.

Recovery of the downstream sequence of truncated *xabC* by IPCR

Genomic DNA of *X. albilineans* LS155 was digested with *Nco*I. Following phenol/chloroform extraction and ethanol precipitation, the digested DNA was self-ligated at a concentration of 0.5 µg/mL. The ligated DNA was precipitated with ethanol and resuspended in sterile H₂O to a concentration of 20 ng/µL as template for IPCR. Sequence of the 16.5 kb *Eco*RI fragment including the 5' region of *xabC* was used to design primers (IF: 5'-AAGCGTCGACATAGCAGCAG-3' and IR: 5'-CGGCAACGCATTCGACCTCG-3') for IPCR-amplification of the sequence downstream of the *Eco*RI site of truncated *xabC* gene.

IPCR was performed in a volume of 50 µL with 50 ng of template DNA, 0.4 ng/µL of each of primer, 0.2 mM of each of dNTP, 1.8 mM Mg²⁺, and 1 unit of elongase enzyme mix with proof-reading activity (Life Technologies). A 10 min initial denaturation step at 94° C was followed by 35 thermal cycles of denaturation at 94° C for 1 min, annealing at 55° C for 1 min, and extension at 72° C for 1 min per 1 kb of expected amplification product. The IPCR product was cloned into pZerO-2 to give pZIXC. Clones of construct pZIXC from three independent PCR reactions were sequenced to rule out the possibility of PCR-generated errors.

Insertional mutagenesis

An internal 625 bp *Cla*I-*Eco*RI fragment of *xabC* (Figure 13) was firstly cloned into *Cla*I/*Eco*RI-digested pBluescript II SK to provide a *Kpn*I restriction site, then subcloned into *Eco*RI/*Kpn*I-cleaved pJP5603 to yield pJP-BEC. The inserts in pBluescript II SK intermediates (pBEC) were sequenced to confirm the expected clones.

The suicide construct pJP-BEC was transferred from the mobilising strain *E. coli* S17-1 (λpir) into *X. albilineans* LS155. Exconjugant colonies were selected on SP agar containing kanamycin and ampicillin. Insertional disruption in *xabC* or *thp* was verified by PCR using primers flanking the expected integration site of pJP-BEC or pJP-BAS and extension at 72° C for 1 min as previously described (Zhang and Birch, 1997b). The effect on albicidin biosynthesis was determined using the microbial plate assay. Representative (Tox⁻) insertional mutants in *xabC* (LS-JP1) and *thp* (LS-JP2) were retained for further analysis.

Construction of expression vectors

The coding region of the *xabC* gene was amplified from *X. albilineans* LS155 chromosomal DNA by PCR. Primer A3F (5'-CGGGATCCCATGGATTGAGCGTTACC-3') contained a BamHI restriction site (underlined) for insertion of the amplified gene into the correct reading frame of *lacZ* in pLAFR3. Primer A3R (5'-CCCAAGCTTTTCATTATGGGGCCCTCTTGC-3') introduced a HindIII restriction site (underlined). The amplified DNA was digested with BamHI and HindIII, and ligated with BamHI/HindIII-digested pLAFR3 to result in pLXABC. *X. albilineans* Tox⁻ mutant LS157 contains a single Tn5 insertion, in a 4.1 kb *Cla*I restriction fragment or a 16.5 kb *Eco*RI restriction fragment (Figure 12). Selection for kanamycin resistance, following shotgun cloning of *Cla*I restriction fragments of LS157 DNA into pBluescript II SK, yielded clone pBC157. Sequences flanking the Tn5 insertion in LS157 DNA were amplified by inverse PCR, and cloned into pZER0-2, producing pZIL and pZIR. The double-strand sequence of the 16,511 bp *Eco*RI genomic fragment in pLXABB was obtained by a primer-walking approach, using subclones pBC157, pZIL, pZIR, pSEBL, and pSEBR. The Tn5 insertion in the genome of LS157 is accompanied by 9-bp perfect repeat sequence (GTCCTGAAG), commencing at 2490 bp in GenBank accession no. AF239749.

Genetic complementation of albicidin biosynthesis

DNA transfer between *E. coli* donor (JM109 pLAFR3 ± insert) and *X. albilineans* recipient (LS-JP1 or LS-JP2), was accomplished by triparental transconjugation with helper strain pRK2013. Mid-log-phase cultures of the recipient were spotted onto agar plates containing YEB medium with no antibiotics (20 µL per spot). After the liquid was absorbed by the agar, 20 µL of mid-log-phase culture of the helper was added to each spot. The liquid was again allowed to absorb, and 20 µL of mid-log-phase culture of the donor was added to each spot. After incubation of the mating plates overnight at 28° C, transconjugants were selected on SP plates supplemented with ampicillin, and tetracycline or spectinomycin.

Transconjugants were tested for albicidin production using the microbial plate bioassay. The constructs pLXABB, pLXABC were designed to test complementation in trans. However, complementation could also occur in *cis*, by homologous recombination between the complementing construct and the insertionally mutated chromosomal gene. To

exclude this possibility, the retention of the insertion in *xabC* was confirmed by PCR, using primers from *aphA* (in the insertion) and *xabB* (adjoining *xabC* in the chromosome).

Results and Discussion

Cloning and sequencing of the full-length *xabC* gene

- 5 Downstream by 45 bp from the TAG stop codon of *xabB* is the start of an ORF (designated *xabC*) in the same orientation. The 639-bp sequence downstream of the *EcoRI* site of the truncated *xabC* was amplified from wt *X. albilineans* LS155 using IPCR. The double-strand nucleotide sequence of 1515 bp from the stop codon of *xabB* to the *NcoI* site downstream of *xabC* (Figure 13) is deposited in GenBank under accession no. AF239750.
- 10 The *xabC* ORF encodes a protein of 343 aa (Mr 37,704). One TCTG-like sequence (TGTG) and one typical TCTG box characteristic of many factor independent termination sites (Brendel and Trifonov, 1984) occur downstream of the termination codon (TAA) of *xabC* (Fig. 2). However, the other features typical of such terminators (a region of GC rich dyad symmetry, followed by a run of consecutive thymine residues) are not present within
- 15 435 bp downstream of the *xabC* stop codon.

XabC is similar to O-methyltransferases

- The deduced product of *xabC* shows 22-30% overall identity and 52-60% overall similarity to a family of methyltransferases that utilise S-adenosyl-methionine (SAM) as a co-substrate for O-methylation of small molecules (Ingrosso *et al.*, 1989; Haydock *et al.*, 1991; Kagan and Clarke, 1994). These enzymes include tetracenomycin polyketide C-8 O-methyltransferase (TcmO, P39896) and C-3 O-methyltransferase (TcmN, P16559) of
- 20 *Streptomyces glaucescens*, hydroxyneurosporene-O-methyltransferase (P17061) of *Rhodobacterium capsulatus*, and hydroxyindole-O-methyltransferases of rat pineal and retina (O09179) and chicken pineal gland (Q92056). Three highly conserved motifs in
- 25 SAM-dependent methyltransferases are also present in XabC as shown in Figures 13 and 14. The crystal structure analysis for the methyltransferase-SAM complex (Schlukebieer *et al.*, 1995) provides firm structural evidence for the role of motif I in SAM binding.

Insertional mutagenesis of *xabC* blocks albicidin biosynthesis

Insertional mutation in *xabC* was accomplished using suicide-vector pJP-BEC and confirmed by PCR. Six out of eight tested transconjugants were verified by PCR to contain insertional mutations in *xabC*. Albicidin production was undetectable in these insertional mutants, compared to wt *X. albilineans* LS155 control. The other transconjugants may result from integration of the vector at other genomic locations by illegitimate recombinations as reported previously (Penfold and Pemberton, 1992).

Complementation test

Introduction of the *xabC* gene in pLXABC or the truncated *xabC* gene in pLXABB into insertional mutant LS-JP2 restored albicidin production to the level of the wt parental strain LS155. This indicates that *xabC* is essential for albicidin production in *X. albilineans*. The truncated *xabC* in pLXABB (SEQ ID NO: 106) encodes 277 residues (SEQ ID NO: 107), including all of the three conserved motifs of SAM-methyltransferases, and appears fully functional by complementation. The continued presence of an insertion in the chromosomal locus was confirmed by PCR. Thus, complementation was operating in *trans*. This also indicates that no other cistron downstream of *xabC* is required for albicidin production, because insertional mutagenesis typically causes polar mutations affecting all downstream cistrons in an operon (De Bruijn and Lupski, 1989).

Enhanced expression of *xabC* results in increased production of albicidins

Derivatives of *X. albilineans* strain LS155, in which an *xabC* gene, or fragment thereof, was introduced in *trans*, were tested for production of albicidin using the bioassay described above. The results, presented in Figure 15, show that expression of *xabC* cloned into pLAFR3 in derivatives of *X. albilineans* strain LS155 complements an insertional mutation in the chromosomal *xabC*, and also enhances albicidin production early in the production phase. Expression of the first part of the *xabB* operon, including the full-length *xabB* and a truncated but functional *xabC*, further enhances albicidin production.

- 92 -

The disclosure of every patent, patent application, and publication cited herein is hereby incorporated herein by reference in its entirety.

The citation of any reference herein should not be construed as an admission that such reference is available as "Prior Art" to the instant application

- 5 Throughout the specification the aim has been to describe the preferred embodiments of the invention without limiting the invention to any one embodiment or specific collection of features. Those of skill in the art will therefore appreciate that, in light of the instant disclosure, various modifications and changes can be made in the particular embodiments exemplified without departing from the scope of the present
- 10 invention. All such modifications and changes are intended to be included within the scope of the appended claims.

TABLES

TABLE 1

Bacterial strains, and plasmids for Example 1

<i>Strain or plasmids</i>	<i>Relevant characteristics</i>	<i>Reference or source</i>
Strains		
<i>E. coli</i>		
DH5 α	Φ 80dlacZAM15, Δ (lacZYA-argF	Promega
JM109	[F', lacI ⁺ ZAM15], Δ (lac-proAB	Promega
TOP10	F', Δ (mrr-hsdRMS-mcrBC), Δ (are-leu)7697, Δ lacX74	Invitrogen
<i>X. albilineans</i>		
Xa13	Wild-type albicidin producer from sugarcane (Queensland), Ap ^r	Inventor's laboratory
LS155	Wild-type albicidin producer from sugarcane (Queensland), Ap ^r	Wall and Birch (1997)
LS157	LS155::Tn5, albicidin deficient (Tox ⁻), Km ^r St ^r Ap ^r	Wall and Birch (1997)
Plasmids		
pBluescript II SK	ColE1 origin, E. coli cloning vector, Ap ^r	Stratagene
pZerO-2	ColE1 origin, E. coli cloning vector, Km ^r	Invitrogen
pRK2013	ColE1 origin, IncP, Tra ⁺ , helper plasmid, Km ^r	Ditta et al ('980)
pLAFR3	RK2 origin, Tra ⁻ , Mob ⁺ , broad host-range cosmid, Tc ^r	Stachelhaus et al. (1987)
pRG960sd	ColE1 origin, broad host-range plasmid, contains promoterless uidA with start codon and Shine-Dalgarno sequence, Sm ^r Sp ^r	Van den Edde et al. (1992)
pBC157	9.9-kb ClaI fragment carrying Tn5 and flanking sequences from LS157, in pBluescript II SK, Km ^r Ap ^r	This study
pZIL	1.4-kb fragment, inverse PCR amplified from LS157 in pZerO-2, Km ^r	This study
pZIR	6.0-kb fragment, inverse PCR amplified from LS157 in pZerO-2, Km ^r	This study
pZTI	0.9-kb fragment, PCR amplified from LS157 in pZerO-2, Km ^r	This study
pXABB	16.5-kb EcoRI fragment from Xa13 in pBluescript II SK, Ap ^r	This study
pSEBL	7.9-kb EcoRI-SpeI fragment from pXABB in pBluescript II SK, Ap ^r	This study
pSEBR	8.6-kb EcoRI-SpeI fragment from pXABB in pBluescript II SK, Ap ^r	This study

- 94 -

<i>Strain or plasmids</i>	<i>Relevant characteristics</i>	<i>Reference or source</i>
pLXABB1	16.5-kb EcoRI fragment from pXABB in pLAFR3 (xabB in the same direction as lac), Tc ^r	This study
pLXABB2	16.5-kb EcoRI fragment from pXABB in pLAFR3 (xabB in the opposite direction to lac), Tc ^r	This study
pRG960p1	206-bp BamHI-XmaI fragment in pRG960sd, Sm ^r Sp ^r	This study
pRG960p2	690-bp BamHI-XmaI fragment in pRG960sd, Sm ^r Sp ^r	This study

TABLE 2

Comparison of conserved sequences in peptide synthetases and XabB

Domain	Core	Sequence conserved in peptide synthetases ^a	Sequence in XabB	Position in Xab (aa)
Adenylation	A1	L (T/S) YxEL	WSYAQL	3806-3811
	A2	LKAGxAYL (V/L) P (L/I) D	FKAGACYVPID	3851-3861
	A3	LAYxYTSG (S/T) TGxPKG	LACVMVTSGSTGRPKG	3917-3932
	A4	FDxS	FAVS	3967-3970
	A5	NxYGPTe	NNYGCTE	4063-4069
	A6	GELxLxGxG (V/L) ARGYL	GELHVHVSVMARGYW	4114-4128
	A7	Y (R/K) TGDL	YKTGDM	4152-4157
	A8	GRxDxQVKIRGxRIELGEIE	GRQDFEVKVRGHRVDTRQVE	4170-4189
	A9	LPxYM (I/V) P	LPTYMLP	4239-4245
	A10	NGK (V/L) DR	NGKDDR	4259-4264
Peptidyl carrier protein	PCP	DxFFxLGG (H/D) S (L/I)	DNFFALGGHSL MDFFAVGGHSV	4306-4316 3261-3271
Condensation	C1	SxAQxR (L/M) (W/Y) xL	TYAQERLWLV SLFQERLWLV	3333-3342 4374-4383
	C2	RHExLRTxF	RHEVLRTRF RHEILRTRF	3381-3389 4421-4429
	C3	MHHxISDG (W/V) S	IHHIISDGWS MHHLIYDAWS	3456-3465 4498-4507
	C4	YxD (F/Y) AVW	YADYALW YADYAIW	3495-3501 4538-4544
	C5	(I/V) GxFVNT (Q/L) (C/A) xR	IGFFINILPLR IGFFINILPLR	3606-3617 4649-4659
	C6	(H/N) QD (Y/V) PFE	HQSVPFEE NQALPFE	3641-3647 4685-4691
	C7	RDxSRNPL	RDSSQIPL RDTSRIPL	3658-3665 4701-4708

^aSourced from reference (Marahiel et al., 1997).

TABLE 3

Bacterial strains, and plasmids for Example 2

<i>Strain or plasmids</i>	<i>Relevant characteristics</i>	<i>Reference or source</i>
Strains		
<i>E. coli</i>		
DH5 α	Φ 80dlacZAM15, recA1, endA1, gyrA96, thi-1, hsdR17(r _k ⁻ , m _k ⁺) supE44, relA1, deoR, Δ (lacZYA-argF)U169	Promega
JM109	[F ⁺ , traD36, proAB, lacI ^q ZAM15], recA1, endA1, gyrA96, thi hsdR17(r _k ⁻ , m _k ⁺), supE44, relA1, Δ (lac-proAB)	Promega
<i>X. albilineans</i>		
Xa13	Wild-type albicidin producer from sugarcane (Queensland), Ap ^r	This laboratory
LS155	Wild-type albicidin producer from sugarcane (Queensland), Ap ^r	Wall & Birch (1997)
LS156	LS155::Tn5, albicidin deficient (Tox ⁻), Km ^r S ^r Ap ^r	Wall & Birch (1997)
Plasmids		
pBluescript II SK	ColE1 origin, <i>E. coli</i> cloning vector, Ap ^r	Stratagene
pGEM-T	ColE1 origin, <i>E. coli</i> TA-cloning vector, Ap ^r	Promega
pRK2013	ColE1 origin, IncP, Tra ⁺ , helper plasmid, Km ^r	Ditta et al (1980)
pLAFR3	RK2 origin, Tra ⁻ , Mob ⁺ , broad host-range cosmid, Tc ^r	Staskawicz et al. (1987)
pBEA1	8.8-kb <i>Eco</i> RI fragment carrying Tn5 and flanking sequences from LS156, in pBluescript II SK, Km ^r Ap ^r	This study
pCEA1	1766-bp <i>Eco</i> RI- <i>Cl</i> aI fragment from pBEA1 in pBluescript II SK, Ap ^r	This study
pPEA1	697-bp <i>Eco</i> RI- <i>P</i> stI fragment from pBEA1 in pBluescript II SK, Ap ^r	This study
pGTA1	2.1-kb fragment, PCR amplified from LS155 in pGEM-T, Ap ^r	This study
pLXABA	834-bp <i>Eco</i> RI- <i>B</i> amHI fragment (xabA ORF) from pGTA1 in pLAFR3, Tc ^r	This study
pLENTD	630-bp <i>Eco</i> RI- <i>H</i> indIII fragment (entD ORF) from DH5 α in pLAFR3, Tc ^r	This study

TABLE 4

Similarity between XabA and other PPTases involved in antibiotic and fatty acid biosynthesis in bacteria

Pathway	Protein	Organism	Specificity (A/P)†	Domain I	Domain II	Homology (ID/SIM)
Albicidin	XabA	X.albilineans	?	GVGIDLERP-- (X) 39--	PSAKESLPKA Y	-
Enterobactin	EntD	E.coli	P†	PIGIDIEEI-- (X) 36--	PSAKESAPKASE	35.9/56.5
		S.flexneri	?	PIGV DIEEI-- (X) 36--	PSAKESAPKAS?	35.0/55.6
		S.typhimurium	?	RIGIDIEKI-- (X) 35--	PSAKESVYKAPQ	35.9/62.0
		S.austin	?	RVGV DIEKI-- (X) 35--	PSAKESVYKAPQ	36.1/61.1
Mycobactin	PptT	M.tuberculosis	P	SVGIDAEHP-- (X) 34--	PCAKEATYKAWF	30.5/55.5
Surfactin	Sfp	B.subtilis	A/P†	PIGIDIEKT-- (X) 35--	WSMKESFIKQES	24.8/48.5
	Psf-1	B.pumilus	?	PVGIDIEEI-- (X) 35--	WSMKESFIKLTG	19.8/47.6
Gramicidin	Gsp	B.brevis	P†	PVGIDIERI-- (X) 35--	WTIKESYIKAIG	20.8/42.0
Iturin A	Lpa-14	B.subtilis	?	PIGIDIEKM-- (X) 35--	WSMKESFIKQAG	20.0/43.4
Fatty acids	HI0152	H. influenzae	?	AVGIDIEFP-- (X) 34--	WCLREAVLKSQS	19.7/45.7
	AcpS	E. coli	A†	GLGTDIVEI-- (X) 40--	FAVKEAAKAPG	16.5/38.8
		M.tuberculosis	A	GVGIDL VSI-- (X) 41--	WAAKEAVIKAWS	25.7/47.6
		B. subtilis	?	GIGLDITEL-- (X) 41--	FAAKEAFSKAPG	25.5/46.2
PPTase domain*				(V/I)G(I/V)D	(E/W) (S/C/T) XKE (S/A) XXX	

TABLE 5

Bacterial strains, and plasmids for Example 3

Strain or plasmids	Characteristics	Reference or source
<u>Strain</u>		
<i>E. coli</i>		
DH5 α	Φ 80dlacZ Δ M15, Δ (lacZYA-argF)U169	Promega
JM109	[F', lacI ^q Z Δ M15], Δ (lac-proAB)	Promega
TOP10	F, Δ (mrr-hsdRMS-mcrBC), Δ (are-leu)7697, Δ lacX74	Invitrogen
S17-1 λ pir	S17-1 lysogenized with λ pir	Penfold and Pemberton (1992)
<u>X. albilineans</u>		
Xa13	wt albicidin producer from sugarcane (Queensland), Ap ^r	Our laboratory
LS155	wt albicidin producer from sugarcane (Queensland), Ap ^r	Wall and Birch (1997)
LS157	xabB::Tn5, albicidin deficient (Tox ⁻), Km ^r St Ap ^r	Wall and Birch (1997)
LS-JP1	thp::pJP-BAS, albicidin deficient (Tox ⁻), Km ^r Ap ^r	This work
LS-JP2	xabC::pJP-BEC, albicidin deficient (Tox ⁻), Km ^r Ap ^r	This work
<u>Plasmids</u>		
pBluescript II SK	ColE1 origin, E. coli cloning vector, Ap ^r	Stratagene
pZER0-2	ColE1 origin, E. coli cloning vector, Km ^r	Invitrogen
pRK2013	ColE1 origin, IncP, Tra ⁺ , helper plasmid, Km ^r	Ditta <i>et al.</i> (1980)
pLAFR3	RK2 origin, Tra ⁻ , Mob ⁺ , broad host-range cosmid, Tc ^r	Staskawicz <i>et al.</i> (1987)
pJP5603	Bacterial suicide vector, Km ^r	Penfold and Pemberton (1991)
pZIXC	1 kb IPCR product in pZER0-2, Km ^r	This work
pBAS	278 bp ApaI-SalI fragment of thp in pBluescript II SK, Ap ^r	This work

<i>Strain or plasmids</i>	<i>Characteristics</i>	<i>Reference or source</i>
pJP-BAS	284 bp SalI-KpnI fragment from pBAS in pJP5606, Km ^r	This work
pBEC	625 bp ClaI-EcoRI fragment of xabC in pBluescript II SK, Ap ^r	This work
pJP-BEC	655 bp EcoRI-KpnI fragment from pBEC in pJP5603, Km ^r	This work
pLTHP	1226 bp EcoRI-BamHI fragment from pLXABB in pLAFR3, Tc ^r	This work
pLXABC	1029 bp xabC ORF amplified from LS155 in pLAFR3, Tc ^r	This work
pLXABB	16.5 kb EcoRI fragment from Xa13 in pLAFR3, Tc ^r	This work

BIBLIOGRAPHY

- Albertini, A.M., Caramori, T., Scoffone, F., Scotti, C. and Galizzi, A. (1995) Sequence around the 159^o region of the *Bacillus subtilis* genome: the *pksX* locus spans 33.6 kb. *Microbiol* 141: 299-309.
- 5 Aparicio, J.F., Molnar, I., Schwecke, T., Konig, A., Haydock, S.F., Khaw, L.E., *et al.* (1996) Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* 169: 9-16.
- Bender, C. L., Alarcon-Chaidez, F. & Gross, D. C. (1999). *Pseudomonas syringae* 10 phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* 63: 266-292.
- Birch, R.G., and Patil, S.S. (1983) The relation of blocked chloroplast differentiation to sugarcane leaf scald disease. *Phytopathology* 73: 1368-1374.
- Birch, R. G. & Patil, S. S. (1985a). *Antibiotic and process for the production thereof.* 15 U.S. patent 4525354, 25 June 1985.
- Birch, R.G., and Patil, S.S. (1985b) Preliminary characterization of an antibiotic produced by *Xanthomonas albilineans* which inhibits DNA synthesis in *Escherichia coli*. *J Gen Microbiol* 131: 1069-1075.
- Birch, R.G., and Patil, S.S. (1987a) Correlation between albicidin production and 20 chlorosis induction by *Xanthomonas albilineans*, the sugarcane leaf scald pathogen. *Physiol Mol Plant Pathol* 30: 199-206.
- Birch, R.G., and Patil, S.S. (1987b) Evidence that an albicidin-like phytotoxin induces chlorosis in sugarcane leaf scald disease by blocking plastid DNA replication. *Physiol Mol Plant Pathol* 30: 207-214.
- 25 Borchert, S., Stachelhaus, T. & Marahiel, M. A. (1994). Induction of surfactin production in *Bacillus subtilis* by *gsp*, a gene located upstream of the gramicidin 3 operon in *Bacillus brevis*. *J Bacteriol* 8: 2458-2462.

- 101 -

- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254.
- 5 Brendel, V. & Trifonov, E. N. (1984). A computer algorithm for testing potential prokaryotic terminator. *Nucleic Acids Res* 12: 4411-4427.
- Cane, D.E., Walsh, C.T., and Khosla, C. (1998) Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* 282: 63-68.
- Coderre, P. E. & Earhart, C. F. (1989). The *entD* gene of the *Escherichia coli* K12 enterobactin gene cluster. *J. Gen Microbiol* 135: 3043-3055.
- 10 Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* 16: 4174-4183.
- Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., and Leadlay P.F. (1990) An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* 348(6297): 176-178.
- 15 De Bruijn, F. J. & Lupski, J. R. (1984). The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids - a review. *Gene* 27: 131-149.
- Ditta G., Stanfield, S., Corbin, D., and Helinski, D.R. (1980) Broad host range DNA cloning system for Gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc Natl Acad Sci USA* 77: 7347-7351.
- 20 Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* 252: 675-679.
- 25 Donadio S., and Katz L. (1992) Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* 111(1): 51-60.

Donadio, S., Staver, M.J., 1993. IS136, an insertion element in the erythromycin gene cluster of *Saccharopolyspora erythraea*. *Gene* 126: 147-151.

Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., *et al.* (1999) The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a
5 multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc Natl Acad Sci USA* 96: 13294-13299.

Gehring, A. M., Bradley, K. A. & Walsh, C. T. (1997a). Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is
10 phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* 36: 8495-8503.

Gehring, A. M., Lambalot, R. H., Vogel, K. W., Drueckhammer, D. G. & Walsh, C. T. (1997b). Ability of *Streptomyces* spp. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for *E. coli* holo-ACP synthase. *Chem Biol* 4:17-24.

Gehring, A.M., DeMoll, E., Fetherston, J.D., Mori, I., Mayhew, G.F., Blattner, F.R., *et al.* (1998) Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*. *Chem Biol* 5: 573-586.
15

Gross, D.C., 1991. Molecular and genetic analysis of toxin production by pathovars of *Pseudomonas syringae*. *Ann. Rev. Phytopath.* 29: 247-278.

Grossman, T. H., Tuckman, M., Ellestad, S. & Osburne, M. S. (1993). Isolation
20 and characterization of *Bacillus subtilis* genes involved in siderophore biosynthesis: relationship between *B. subtilis* *sfp*^o and *Escherichia coli* *entD* genes. *J Bacterio*. 19: 6203-6211.

Guenzi, E., Galli, G., Grgurina, I., Gross, D.C., and Grandi, G. (1998) Characterization of the syringomycin synthetase gene cluster: a link between prokaryotic
25 and eukaryotic peptide synthetases. *J Biol Chem* 273: 32857-32863.

Haydock, S.F., Dowson, J.A., Dhillon, N., Robert, G.A., Cortes, J., Leadlay, P.F., 1991. Cloning and sequencing analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea*: sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. *Mol. Gen. Genet.* 230: 120-128.

Hopwood, D.A. and Khosla, C., Secondary Metabolites: Their Function and Evolution (1992) Wiley Chichester (Ciba Foundation Symposium 171) pp. 88-112

Hopwood, D.A., and Sherman, D.H. (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* 24: 37-66.

- 5 Hopwood, D.A. (1997) Genetic contributions to understanding polyketide synthases. *Chem Rev* 97: 2465-2497.

Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998) Crystal structure of β -ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J* 17: 1183-1191.

- 10 Ingrosso, D., Fowler, A. V., Bleibaum, J., Clarke, S., 1989. Sequence of the D-aspartyl L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs from protein, DNA, RNA and small molecule S-adenosylmethionine dependent methyltransferase. *J. Biol. Chem.* 264: 20131-20139.

- 15 Jefferson, R.A. (1987) Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol Biol Rep* 5: 387-405.

Kagan, R.M., Clarke, S., 1994. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferase suggests a common structure for these enzymes. *Arch. Biochem. Biophys.* 310: 417-427.

- 20 Kealey, J. T., Liu, L., Santi, D. V., Betlach, M. C. & Barr, P. J. (1998). Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc Natl Acad Sci USA* 95: 505-509.

Keating, T.A., and Walsh, C.T. (1999) Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Curr Opin Chem Biol* 3: 598-606.

- 25 Konz, D., and Marahiel, M.A. (1999) How do peptide synthetases generate structural diversity? *Chem Biol* 6: R39-R48.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* 234: 187-208.

- 104 -

- Lambalot, R. H. & Walsh, C. T. (1995). Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *J Biol Chem* 270: 24658-24661.
- Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., et al. (1996) A new enzyme superfamily – the phosphopantetheinyl transferases. *Chem Biol* 3: 923-936.
- MacNeil, D.J., Occi, J.L., Gewain, K.M., MacNeil, T., Gibbons, R.H., Ruby, C.L., and Danis, A.J. (1992) Complex organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase. *Gene* 115: 119-125.
- Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* 97: 2651-2673.
- Martin, J.F., Liras, P., 1989. Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Ann. Rev. Microbiol* 43: 173-206.
- McCarthy, J.E.G., and Gualerzi, C. (1990) Translational control of prokaryotic gene expression. *Trends Genet* 6: 78-85.
- Mofid, M. R., Marahiel, M. A., Ficner, R. & Reuter, K. (1999). Crystallization and preliminary crystallographic studies of Sfp: a phosphopantetheinyl transferase of modular peptide synthetases. *Acta Cryst D* 55: 1098-1100.
- Motamedi, H., and Shafiee, A. (1998) The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur J Biochem* 256: 528-534.
- Nakano, M. M., Corbell, N., Besson, J. & Zuber, P. (1992). Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. *Mol Gen Genet* 232: 313-321.
- Paitan, Y., Alon, G., Orr, E., Ron, E.Z., and Rosenberg, E. (1999) The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *J Mol Biol* 286: 465-474.

- Penfold, R.J., Pemberton, J.M., 1992. An improved suicide vector for construction of chromosomal insertion mutations in bacteria. *Gene* 118: 145-146.
- Platt, T. (1986). Transcription termination and the regulation of gene expression. *Ann Rev Biochem* 55: 339-72.
- 5 Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T. (1998a). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* 37: 1585-1595.
- Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H. & Walsh, C. T. (1998b). Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic
10 enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol* 5: 631-645.
- Rangaswamy, V., Mitchell, R., Ullrich, M., Bender, C., 1998. Analysis of genes involved in biosynthesis of coronafacic acid, the polyketide component of the phytotoxin coronatine. *J. Bacteriol.* 180: 3330-3338.
- 15 Reimman, C., Moore, R., Little, S., Savioz, A., Willetts, N.S., Haas, D., 1989. Genetic structure, function and regulation of the transposable element IS21. *Mol. Gen. Genet.* 215: 416-424.
- Ricaud, C., and Ryan, C.C. (1989) Leaf scald, In *Disease of Sugar Cane: major disease*. Ricaud, C., Egan, B.T., Gillaspie Jr., A.G., and Hughes, C.G. (eds). Amsterdam:
20 Elsevier Science Publisher B. V., pp. 39-53.
- Rott, P.C., Costet, L., Davis, M.J., Frutos, R., and Gabriel, D.W. (1996) At least two separate gene clusters are involved in albicidin production by *Xanthomonas albilineans*. *J Bacteriol* 178: 4590-4596.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A
25 Laboratory Manual*, 2nd edn. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.
- Sanger, F., Nickles, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467.

- 106 -

Schluckebier, G., O'Gara, M., Sanger, W., Cheng, X.D., 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**: 15-20.

Schmid, S., Seitz, T., Haas, D., 1998. Cointegrase, a naturally occurring truncated form of IS21 transposase, catalyzes replicon fusion rather than simple insertion of IS21. *J. Mol. Biol.* **282**: 571-583.

Scrutton, N.S., Berry, A., and Perham, R.N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* **343**: 38-43.

Solinas, F., Marconi, A.M., Ruzzi, M., Zennaro, E., 1995. Characterization and sequence of a novel insertion sequence, IS1162, from *Pseudomonas fluorescens*. *Gene* **155**: 77-82.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis - catalytic role of the condensation domain. *J Biol Chem* **273**: 22773-22781.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**: 493-505.

Staskawicz, B., Dahlbeck, D., Keen, N., and Napoli, C. (1987) Molecular characterization of cloned avirulence genes from race 0 and race 1 of *Pseudomonas syringae* pv. *glycinea*. *J Bacteriol* **169**: 5789-5794.

Strohl, W. R. (1992). Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res* **20**: 961-974.

Tang, L., Yoon, Y.J., Choi, C., and Hutchinson, C.R. (1998) Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei*. *Gene* **216**: 255-265.

Turgay, K., Krause, M., and Marahiel, M.A. (1992) Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* **6**: 529-546.

- 107 -

Van den Edde, G., Deblaere, R., Goethals, K., Van Montagu, M., and Hosters, M. (1992) Broad host range and promoter selection vectors for bacteria that interact with plants. *Mol Plant-Microbe Interact* 5: 228-234.

5 Van Larebeke, N., Genetello, C., Hernalsteens, J.P., Depicker, A., Zaenen, I., Messens, E., *et al.* (1977) Transfer of Ti plasmids between *Agrobacterium* strains by mobilization with the conjugative plasmid RP4. *Mol Gen Genet* 152: 119-124.

Wall, M.K., and Birch, R.G. (1997) Genes for albicidin biosynthesis and resistance span at least 69 kb in the genome of *Xanthomonas albilineans*. *Lett Appl Microbiol* 24: 256-260.

10 Walsh, C. T., Gehring, A. M., Weinreb, P. H., Quadri, L. E. & Flugel, R. S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthases. *Curr Opin Chem Biol* 1: 309-315.

Xiao, Y., Lu, Y., Heu, S., and Hutcheson, S.W. (1992) Organization and environmental regulation of the *Pseudomonas syringae* pv. *syringae* 61 *hrp* cluster. *J Bacteriol* 174: 1734-1741.

Zhang, L., and Birch, R.G. (1997) The gene for albicidin detoxification from *Pantoea dispersa* encodes an esterase and attenuates pathogenicity of *Xanthomonas albilineans* to sugarcane. *Proc Natl Acad Sci USA* 94: 9984-9989.

20 Zhang, L., Birch, R. G., 1997b. Mechanism of biocontrol by *Pantoea dispersa* of sugar cane leaf scald disease caused by *Xanthomonas albilineans*. *J. Appl. Microbiol.* 82: 448-454.

Zhang, L., Xu, J., and Birch, R.G. (1998) Factors affecting biosynthesis by *Xanthomonas albilineans* of albicidin antibiotics and phytotoxins. *J Appl Microbiol* 85: 1023-1028.

25 Zhang, L., Xu, J., and Birch, R.G. (1999) Engineered detoxification confers resistance against a pathogenic bacterium. *Nature Biotech* 17: 1021-1024.

CLAIMS

1. An isolated polypeptide comprising at least one domain selected from the group consisting of:
 - (a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or
5 more of SEQ ID NO: 6 and 8, or variants thereof.
 - (b) a β -ketoacyl synthase (KS) domain comprising a sequence set forth in any one or more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;
 - (c) a β -ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID NO: 22, or variants thereof;
 - 10 (d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;
 - (e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.
 - (f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any
15 one or more of SEQ ID NO: 50 and 52, and variants thereof; and
 - (g) a condensation (C) domain comprising a sequence set forth in any one or more of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.
- 20 2. The polypeptide of claim 1, wherein the AL domain comprises each of the sequences set forth in SEQ ID NO: 6 and 8, or variants thereof.
3. The polypeptide of claim 1, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 10, 12 and 14, or variants thereof.
4. The polypeptide of claim 1, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 16, 18 and 20, or variants thereof.
- 25 5. The polypeptide of claim 1, wherein the A domain comprises each of the sequences set forth in SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.
6. The polypeptide of claim 1, wherein the C domain comprises each of the sequences set forth in SEQ ID NO: 54, 56, 58, 60, 62, 64 and 66, or variants thereof.
7. The polypeptide of claim 1, wherein the C domain comprises each of the sequences set
30 forth in SEQ ID NO: 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

- 109 -

8. The polypeptide of claim 1, wherein the domains are arranged in an N- to C-terminal direction as follows: AL-ACP-KS-KR-ACP-ACP-KS-PCP-C-A-PCP-C.
9. The polypeptide of claim 1, comprising the sequence set forth in SEQ ID NO: 2, or biologically active fragment thereof, or variant or derivative of these.
- 5 10. The polypeptide of claim 9, wherein the variant has at least 60% sequence identity to the sequence set forth in SEQ ID NO: 2.
11. The polypeptide of claim 9, wherein the biologically active fragment is at least 6 amino acids in length.
12. An isolated polypeptide comprising at least a biologically active fragment of the
10 sequence set forth in SEQ ID NO: 2 or variant or derivative thereof.
13. The polypeptide of claim 12, wherein the biologically active fragment is at least 6 amino acids in length.
14. The polypeptide of claim 12, wherein the biologically active fragment comprises at least one domain selected from the group consisting of:
- 15 (a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or more of SEQ ID NO: 6 and 8, or variants thereof.
- (b) a β -ketoacyl synthase (KS) domain comprising a sequence set forth in any one or more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;
- (c) a β -ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID
20 NO: 22, or variants thereof;
- (d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;
- (e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.
- 25 (f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 50 and 52, and variants thereof; and
- (g) a condensation (C) domain comprising a sequence set forth in any one or more of SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.

- 110 -

15. The polypeptide of claim 13, wherein the AL domain comprises each of the sequences set forth in SEQ ID NO: 6 and 8, or variants thereof.
16. The polypeptide of claim 13, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 10, 12 and 14, or variants thereof.
- 5 17. The polypeptide of claim 13, wherein the KS domain comprises each of the sequences set forth in SEQ ID NO: 16, 18 and 20, or variants thereof.
18. The polypeptide of claim 13, wherein the A domain comprises each of the sequences set forth in SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.
- 10 19. The polypeptide of claim 13, wherein the C domain comprises each of the sequences set forth in SEQ ID NO: 54, 56, 58, 60, 62, 64 and 66, or variants thereof.
20. The polypeptide of claim 13, wherein the C domain comprises each of the sequences set forth in SEQ ID NO: 68, 70, 72, 74, 76, 78 and 80, or variants thereof.
21. The polypeptide of claim 12, wherein the variant has at least 60% sequence identity to said at least a biologically active fragment.
- 15 22. The polypeptide of claim 12, wherein the variant has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 or 80.
- 20 23. An isolated polypeptide comprising at least biologically active fragment of the sequence set forth in SEQ ID NO: 83, or a variant or derivative thereof.
24. The polypeptide of claim 23, wherein the biologically active fragment comprises at least one of the consensus PPTase sequence motifs set forth in SEQ ID NO: 89 or 93, or variant thereof.
- 25 25. The polypeptide of claim 24, wherein the biologically active fragment comprises both the consensus PPTase sequence motifs set forth in SEQ ID NO: 89 or 93, or variant thereof.

- 111 -

26. The polypeptide of claim 23, wherein the biologically active fragment comprises the intervening sequence between said consensus PPTase sequence motifs, which intervening sequence comprises the sequence set forth in SEQ ID NO: 91, or variant thereof.
27. The polypeptide of claim 23, wherein the biologically active fragment comprises a
5 contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof.
28. The polypeptide of claim 23, wherein the biologically active fragment is at least 6 amino acids in length.
29. The polypeptide of claim 23, wherein the variant has at least 60% sequence identity to
10 the sequence set forth in SEQ ID NO: 83.
30. The polypeptide of claim 23, wherein the variant has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 87, 89, 91 or 93.
31. An isolated polypeptide comprising at least biologically active fragment of the sequence set forth in SEQ ID NO: 95, or a variant or derivative thereof.
- 15 32. The polypeptide of claim 31, wherein the biologically active fragment comprises at least one of the consensus methyltransferase sequence motifs set forth in SEQ ID NO: 99, 101 or 103, or variant thereof.
33. The polypeptide of claim 31, wherein the biologically active fragment comprises all the consensus methyltransferase sequence motifs set forth in SEQ ID NO: 99, 101 and 103, or
20 variant thereof.
34. The polypeptide of claim 31, wherein the biologically active fragment comprises a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 105, or variant thereof.
35. The polypeptide of claim 31, wherein the biologically active fragment comprises a
25 contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 107, or variant thereof.
36. The polypeptide of claim 31, wherein the biologically active fragment is at least 6 amino acids in length.

- 112 -

37. The polypeptide of claim 31, wherein the variant has at least 60% sequence identity to the sequence set forth in SEQ ID NO: 95.
38. The polypeptide of claim 31, wherein the variant has at least 70% sequence identity to any one of the amino acid sequences set forth in SEQ ID NO: 99, 101 or 103.
- 5 39. An isolated polynucleotide comprising a sequence encoding at least one domain selected from the group consisting of:
- (a) an acyl-CoA ligase (AL) domain comprising a sequence set forth in any one or more of SEQ ID NO: 6 and 8, or variants thereof.
 - (b) a β -ketoacyl synthase (KS) domain comprising a sequence set forth in any one or
10 more of SEQ ID NO: 10, 12, 14, 16, 18 and 20, or variants thereof;
 - (c) a β -ketoacyl reductase (KR) domain comprising the sequence set forth SEQ ID NO: 22, or variants thereof;
 - (d) an acyl carrier protein (ACP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 24, 26 and 28, or variants thereof;
 - 15 (e) an adenylation (A) domain comprising a sequence set forth in any one or more of SEQ ID NO: 30, 32, 34, 36, 38, 40, 42, 44, 46 and 48, or variants thereof.
 - (f) a peptidyl carrier protein (PCP) domain comprising a sequence set forth in any one or more of SEQ ID NO: 50 and 52, and variants thereof; and
 - (g) a condensation (C) domain comprising a sequence set forth in any one or more of
20 SEQ ID NO: 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and 80, or variants thereof.
40. The polynucleotide of claim 39, wherein the AL domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 5 or 7, or variants thereof.
41. The polynucleotide of claim 40, wherein the AL domain is encoded by a nucleotide
25 sequence comprising each of the sequences set forth in SEQ ID NO: 5 and 7, or variants thereof.
42. The polynucleotide of claim 39, wherein the KS domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 9, 11, 13, 15, 17 and 19, or variants thereof.

- 113 -

43. The polynucleotide of claim 42, wherein the KS domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 9, 11 and 13, or variants thereof.

5 44. The polynucleotide of claim 42, wherein the KS domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 15, 17 and 19, or variants thereof.

45. The polynucleotide of claim 39, wherein the KR domain is encoded by a nucleotide sequence set forth in SEQ ID NO: 21, or variant thereof.

10 46. The polynucleotide of claim 39, wherein the ACP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 23, 25 and 27, or variants thereof.

47. The polynucleotide of claim 39, wherein the A domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof.

15 48. The polynucleotide of claim 47, wherein the A domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45 and 47, or variants thereof.

49. The polynucleotide of claim 39, wherein the PCP domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 49 and 51, or variants thereof.

20 50. The polynucleotide of claim 39, wherein the C domain is encoded by a nucleotide sequence set forth in any one or more of SEQ ID NO: 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79, or variants thereof.

51. The polynucleotide of claim 50, wherein the C domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 53, 55, 57, 59, 61, 63 and 65, or variants thereof.

25 52. The polynucleotide of claim 50, wherein the C domain is encoded by a nucleotide sequence comprising each of the sequences set forth in SEQ ID NO: 67, 69, 71, 73, 75, 77 and 79, or variants thereof.

- 114 -

53. The polynucleotide of claim 39, comprising the sequence set forth in any one of SEQ ID NO: 1 or 3, or a biologically active fragment thereof at least 18 nucleotides in length, or a polynucleotide variant of these.
54. The polynucleotide of claim 53, wherein the polynucleotide variant has at least 60% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 1 or 3.
55. The polynucleotide of claim 53, wherein the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 1 or 3 under at least low stringency conditions.
56. The polynucleotide of claim 39, wherein the polynucleotide variant comprises a nucleotide sequence encoding at least one said domain.
57. The polynucleotide of claim 56, wherein the nucleotide sequence variant has at least 60% sequence identity to any one or more of the sequences set forth in SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79.
58. The polynucleotide of claim 56, wherein the nucleotide sequence variant is capable of hybridising to any one of the sequences identified by SEQ ID NO: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 and 79 under at least low stringency conditions.
59. An isolated polynucleotide comprising a sequence encoding at least biologically active fragment of the sequence set forth in SEQ ID NO: 83, or a variant or derivative thereof.
60. The polynucleotide of claim 59, comprising the sequence set forth in any one of SEQ ID NO: 82 and 84, or a biologically active fragment thereof, or a polynucleotide variant of these.
61. The polynucleotide of claim 59, comprising a contiguous sequence of nucleotides at least 18 nucleotides in length and contained within the sequence set forth in SEQ ID NO: 86, or variant thereof.
62. The polynucleotide of claim 59, wherein the polynucleotide variant has at least 60% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 82, 84 and 86.

- 115 -

63. The polynucleotide of claim 59, wherein the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 82, 84 and 86 under at least low stringency conditions.

5 64. The polynucleotide of claim 59, wherein the polynucleotide variant comprises a nucleotide sequence encoding at least one PPTase sequence motif selected from SEQ ID NO: 89 and 93, or variant thereof.

10 65. The polynucleotide of claim 64, wherein the polynucleotide variant comprises a nucleotide sequence encoding the intervening sequence between the said consensus PPTase sequence motifs, said nucleotide sequence comprising the sequence set forth in SEQ ID NO: 91.

66. The polynucleotide of claim 59, wherein the polynucleotide variant suitably comprises a nucleotide sequence encoding a contiguous sequence of amino acids contained within the sequence set forth in SEQ ID NO: 87, or variant thereof.

15 67. The polynucleotide of claim 66, wherein the contiguous sequence is encoded by the sequence set forth in SEQ ID NO: 86, or nucleotide sequence variant thereof displaying at 60% identity thereto.

68. The polynucleotide of claim 64, wherein the PPTase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 88 and 92, or nucleotide sequence variant thereof displaying at 60% identity thereto.

20 69. The polynucleotide of claim 65, wherein the intervening sequence is encoded by the nucleotide sequence set forth in SEQ ID NO: 90, or nucleotide sequence variant thereof displaying at 60% identity thereto.

25 70. The polynucleotide of claim 66, wherein the contiguous sequence is encoded by the sequence set forth in SEQ ID NO: 86, or nucleotide sequence variant thereof displaying at 60% capable of hybridising thereto under at least low stringency conditions.

71. The polynucleotide of claim 64, wherein the PPTase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 88 and 92, or nucleotide sequence variant thereof capable of hybridising thereto under at least low stringency conditions.

- 116 -

72. The polynucleotide of claim 65, wherein the intervening sequence is encoded by the nucleotide sequence set forth in SEQ ID NO: 90, or nucleotide sequence variant thereof capable of hybridising thereto under at least low stringency conditions.
73. An isolated polynucleotide comprising a sequence encoding at least biologically active
5 fragment of the sequence set forth in SEQ ID NO: 95, or a variant or derivative thereof.
74. The polynucleotide of claim 73, comprising the sequence set forth in any one of SEQ ID NO: 94 and 96, or a biologically active fragment thereof, or a polynucleotide variant of these.
75. The polynucleotide of claim 73, comprising a contiguous sequence of nucleotides
10 contained within the sequence set forth in SEQ ID NO: 104, or variant thereof.
76. The polynucleotide of claim 73, comprising a contiguous sequence of nucleotides contained within the sequence set forth in SEQ ID NO: 106, or variant thereof.
77. The polynucleotide of claim 73, wherein the polynucleotide variant has at least 60% sequence identity to any one of the polynucleotides set forth in SEQ ID NO: 94, 96, 104
15 and 106.
78. The polynucleotide of claim 73, wherein the polynucleotide variant is capable of hybridising to any one of the polynucleotides identified by SEQ ID NO: 94, 96, 104 and 106 under at least low stringency conditions.
79. The polynucleotide of claim 73, wherein the polynucleotide variant comprises a
20 nucleotide sequence encoding a methyltransferase sequence motif selected from any one or more of SEQ ID NO: 99, 101 and 103, or variant thereof.
80. The polynucleotide of claim 79, wherein the methyltransferase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 98, 100 and 102, or nucleotide sequence variant thereof displaying at least 60%
25 identity thereto.
81. The polynucleotide of claim 79, wherein the methyltransferase sequence motif is encoded by a nucleotide sequence comprising the sequence set forth in any one of SEQ ID NO: 98, 100 and 102, or nucleotide sequence variant thereof capable of hybridising thereto under at least low stringency conditions.

- 117 -

82. An expression vector comprising the polynucleotide of any one of claims 39, 59 or 73, wherein the polynucleotide is operably linked to a regulatory polynucleotide.

83. A host cell containing the expression vector of claim 82.

84. A multiplicity of cell colonies, constituting a library of colonies, wherein each colony
5 of the library contains an expression vector for the production of the polypeptide of claim 1 or claim 12.

85. A method for enhancing the level and/or functional activity of an albicidin, said method comprising:

10 - introducing into an albicidin-producing host cell (1) an agent that modulates the expression of a gene encoding at least a portion of the polypeptide of claim 1 or variant or derivative thereof, or the level and/or functional activity of an expression product of said gene, or (2) a vector from which a polynucleotide encoding at least a portion of the polypeptide of claim 1 or variant or derivative thereof can be translated;

15 - and culturing the host cell for a time and under conditions sufficient to enhance the level and/or functional activity of said albicidin.

86. The method of claim 85, further comprising introducing into said host cell a vector from which a PPTase can be translated.

87. The method of claim 86, wherein the PPTase is selected from EntD or XabA.

20 88. The method of claim 85, further comprising introducing into said host cell a vector from which a methyltransferase can be translated.

89. The method of claim 86, wherein the methyltransferase is XabC.

90. An antigen-binding molecule that is immuno-interactive with the polypeptide of claim 1 or claim 12.

25 91. An antigen-binding molecule that is immuno-interactive with the polypeptide of claim 23.

92. An antigen-binding molecule that is immuno-interactive with the polypeptide of claim 31.

- 118 -

93. A method of preparing a polynucleotide encoding a modified PKS, comprising using a nucleotide sequence encoding the polypeptide of claim 1 or claim 12 as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, deletion, insertion, or replacement.
- 5 94. A method for producing polyketides, comprising expressing the modified albicidin PKS encoding nucleotide sequence produced by the method of claim 93 in a suitable host cell to thereby produce a polyketide different from that produced by said polypeptide.

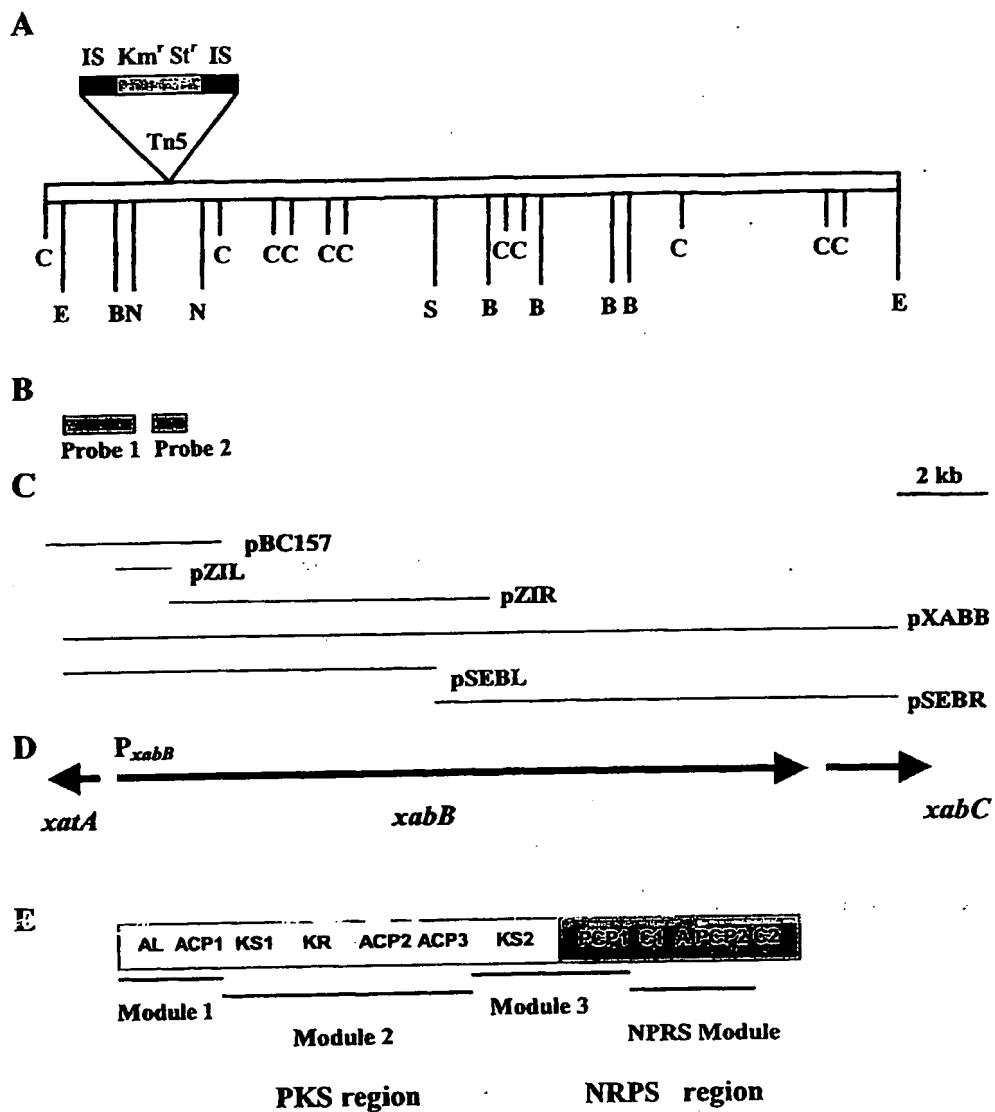


FIGURE 1

PIFI

967 GCGGATGCGC GCACACTGCA GGTCCATCAC GCCACCTCCA GCAGGGTGTC
 CCGCTACGCG CGTGTGACGT CCAGGTAGTG CCGTGGAGGT CGTCCCACAG
 A I R A C Q L D M M RBS

xatA ←

1017 ATACACGGCC AGCGGATGCT GCAGGTTTTT CACTGGCAGG GCCACTGGCT
 TATGTGCCGG TCGCCTACGA CGTCCAAAAG GTGACCGTCC CGGTGACCGA

-35 (*P_{xatB}*)-10 (*P_{xatB}*)

1067 GTCGTAAGGG AAGCGGTGCC TTGAGCGCCG GTGCGGACAG TATAACGACA
 CAGCATTTCCC TTCGCCACGG AACTCGCGGC CACGCCTGTC ATATTGCTGT

-10 (*P_{xatA}*)

1117 CGTTCCTTGG CCAAGCGCAC TGTCGGCAGC GCCTTGCTGA TGCCGCCCAT
 GCAAGGAACC GGTTCGCGTG ACAGCCGTGC CGGAACGACT ACGGCGGGTA

-35 (*P_{xatA}*)

1167 GTAGCCGCGC GCCTGGATCT CGCGTAGTAG CACCACGCTG GCCGGGATCC
 CATCGGCGCG CGGACCTAGA GCGCATCATC GTGGTGCGAC CGGCCCTAGG

PIR

RBS

→ *xatB*

1217 ATCGAGGGCG CGCTTGCCCA ATGCGCTCAT GCAGATAACT CTTGTAGCCG
 TAGCTCCCGC GCGAACGGGT TACGCGAGTA CGTCTATTGA GAACTACGGC

M P N A L M Q I T L V A

FIGURE 2

(i). AL

TSGSSGESKILLSH--GYFRTGDL Xal-XabB(AL)
 TGGTTGVAKGAMLTH--GWMATGDI Hin-LCFA
 TSGSTGTPKAVMLNH--GWFETGDL Bsu-PksJ
 SSGSTGDPKGVMLTH--GWVKTGDL Bsu-MycA(AL)
 SSGTTGLPKGVMLTH--GWLHTGDI Pcr-ComL2
 TSGTTGRPKGVVSAQ--GWYRTGDL Sma-FkbB(AL)
 TSGTTGRPKGVVSTQ--GWFRTGDL Ame-RifA(AL)
 TSGTTGTPKGVLTQ--GWYRTGDL Shy-RapA(AL)

(ii). KS

GPSEVINSACSSSLVAL--VELHGTGTS--ALGHLGAAAG Xal-XabB (KS1)
 GPSLAVDTACSASLTAI--IEAHGTGTVL--NIGHAESAAG Xal-XabB (KS2)
 GPSLFVHTNCSSSLVAL--VEAHGTGTL--NLGHLDTVAG Mxa-Tal
 GPAVTVDTACSSSLVAV--IEAHGTGTL--NIGHLFEAAG Bsu-MycA
 GPAVTVDTACSSSLVAL--VEAHGTGTRL--NIGHAQAAG Scr-EryA1
 GPAMTVDTACSSSLTAL--VEAHGTGTRL--NIGHTQAAG Scr-EryA3
 GPSVLVDTACSSGLTAL--VECHGTGTA--NIGHLEGASG Che-PKS1
 GPSLAVDTACSSSLTAI--LEAHGTGTAL--NIGHCESAAG Bsu-PksM
 GPSVAVDTACSSSLVAI--VEAHGTGTL--NLGHTAAAAG Mtu-PpsA
 GPSLTIDTACSSSLMAL--VEAHGTGTV--NMGHPEPASG Chick-FAS
 GPSIALDTACSSSLAL--IEAHGTGTV--NMGHPEPASG Rat-FAS

* * *
 (Active site cysteine) (Active site histidine)

(iii). KR

VYVVIGGAGGLGEVLSEHLIRTYD.AQLIWIGR Xal-XabB
 VYVISGGTGALARLFVAEIGKRATRVILVAR Mxa-Tal
 TVLVTGGTGVGQIARWLARRG.APHLLVSR Scr-EryA1
 TVLVTGGTGICAKLARWLARSG.AEHLVLLGR Scr-EryA3
 SYLLVGGVGGGCSATALAMSTRG.ARHLLLINR Che-PKS1
 SYIITGGLGGLGLFFASKLAAAG.CGRIVLTR Mtu-MAS
 SYIITGGLGGFGLAQLWLIERG.AQKLVLTSR Chick-FAS
 SYIITGGLGGFGLARWLVLRG.AQRLVLTSR Rat-FAS

(iv). ACP

CELALDSLQCVR Xal-XabB(ACP1)
 EYGVDSIVAIE Xal-XabB (ACP2)
 ESYGVDSIVIIE Xal-XabB (ACP3)
 IGFGLDSIMLTQ Bsu-MycA
 ERYGIDSIIITQ Mxa-Tal
 AELGVDSLVALE Scr-EryA1
 QDYGIDSLVAVE Che-PKS1
 IEYGLDSLGMLE Mtu-MAS
 ADLGLDSLGMVE Chick-FAS
 ADLGLDSLGMVE Rat-FAS

* (Active site serine)

FIGURE 3

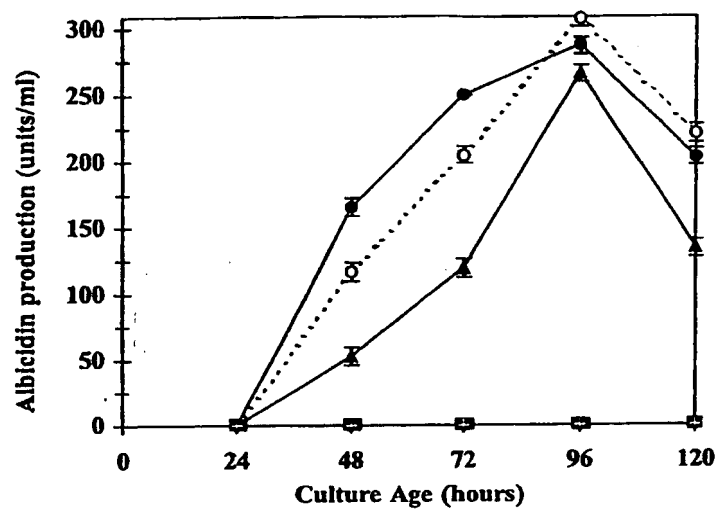


FIGURE 4

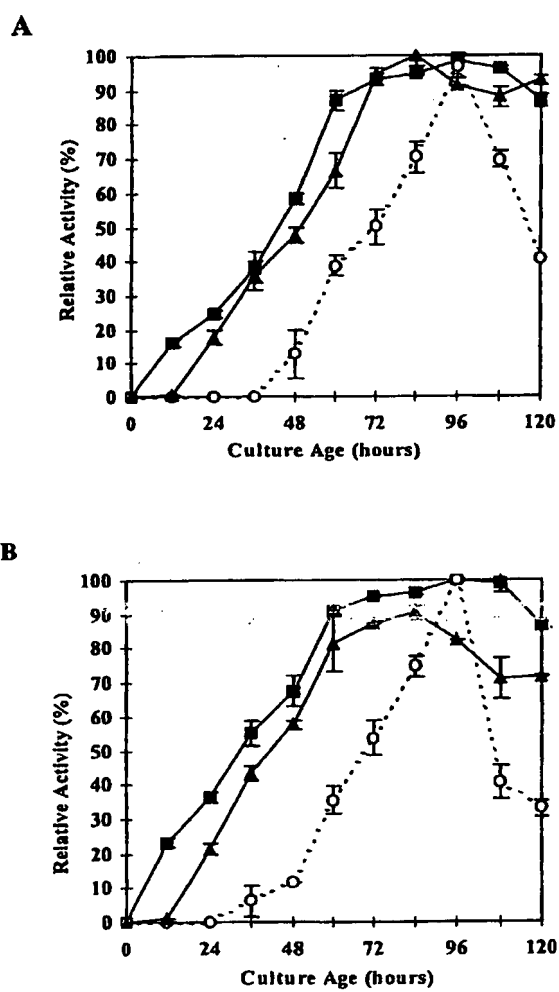


FIGURE 5

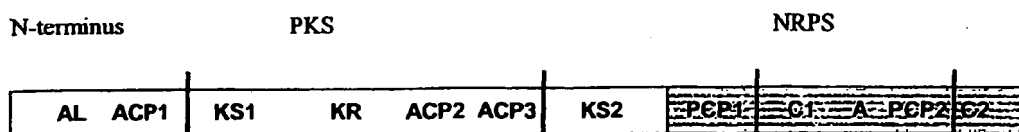
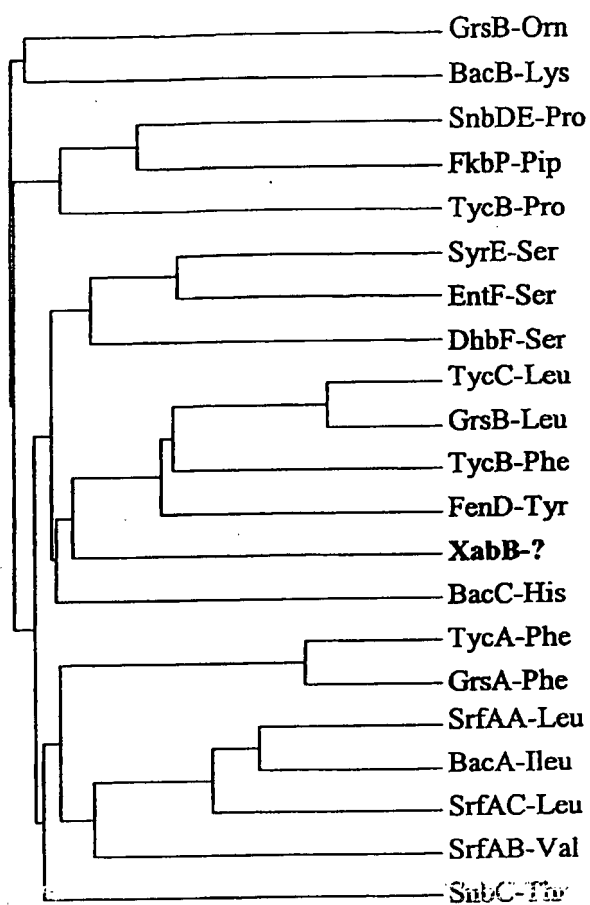
A. *X. albilineans* XabB (4801 aa)B. *B. subtilis* MycA (3971 aa)C. *Yersinia pestis* HMWP1 (3163 aa)D. *M. xanthus* Tal (2392 aa)E. *B. subtilis* PksorFX6 (4447 aa)

FIGURE 6

**FIGURE 7**

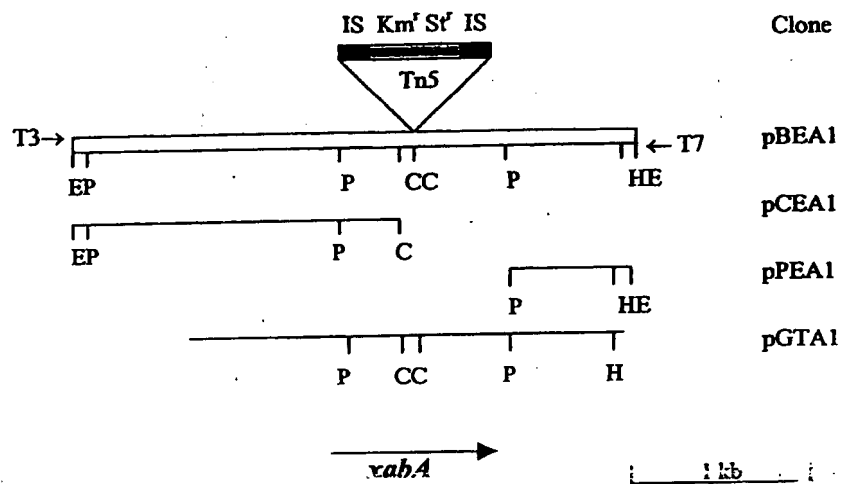
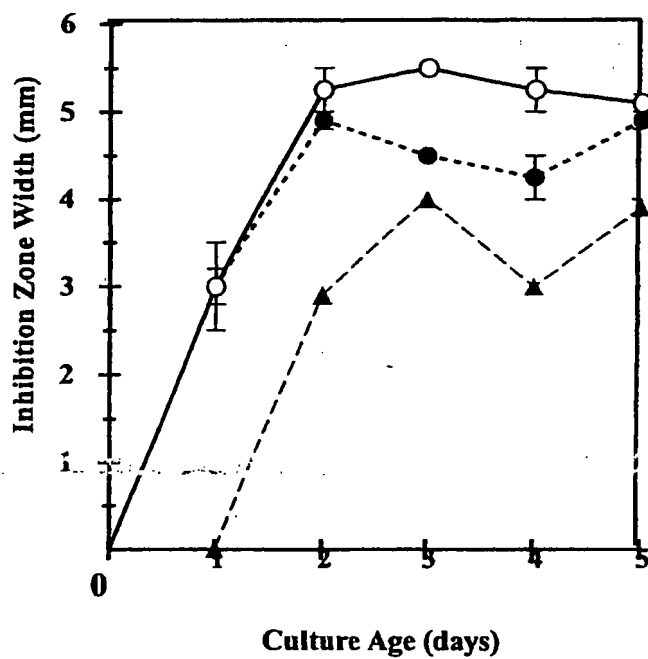
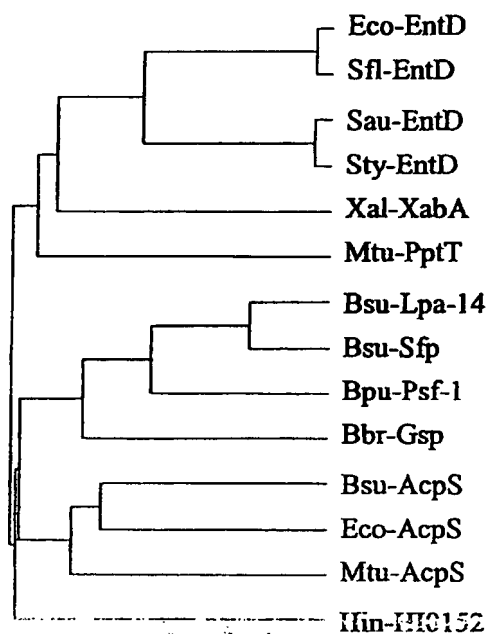


FIGURE 8

1054 TTCCGCCCCGAATAGGCGCAGGAAGCCAATAAGTATGGCAGCGCCCTTGACCAATGACAAGCTCATGCACCCAGGACGCC
1135 GCTCTGCTCCGCGTCCATCGCCATTGCGCCCTCCCGACCCCAAGCATCGACCAAAGGACCGAATGCGGCGGTAGG
1216 CGCGACTCTGCGACTAGCGCAATGTTATCGTGCACATTGACGCCACAGCCCTCAGCGCAACGCAATGCCCAATGCCGT
-10 RBS
1297 ACCGATGCAGGCGCGCGGGGACTCCCGCAGCCGCAAGCGATGAACCCAGGGTTGCCGAGCGTCGGCGGCTTGAGCGCAGG
M P N A V 5
P M Q G A R G L P Q P Q A M N P G L P S V G G L S A G 32
1378 CCAGCCATTGCACTTGTCTAGCACCGGAAGTGCAGGCGCGCGCAGTCCCCACCGCCATCTGCTCGACGACGGCAC
Q P L Q L S L A P E L Q A A A R S A H R H L L D D G T 59
1459 GCGCTTTACCTGCTGGGTTGATACCGCGCAATTGACCCGGGGCTTTCCGGCAATGGCAATCGCCCGCCGACAG
A L Y L L A F D T A Q F D P G A P A A M A I A R P D S 86
1540 CATGCCCGCGAGCGTCCGCAAGCGTCAGGCGGAGTTCTGTTCGGCGCTCGGCCGCGGACTGGCGCTGCAAGAGGTCT
I A R S V R K R Q A E F L P G R L A A R L A L Q E V L 113
1621 GGGACCTGCGCAAGCGCAGGCGAGATATTGCAATCGGCGCGACGCGCGCCCTGCTGGCTGCGGCGAGCTGGGCGCAT
G P A Q A Q A D I A I G A T R A P C W P A G S L G S I 140
1702 TTCCCATTCGAGGACTACGCGCGCCCATCGCCATGGCGCGCGCACCGCCACGGCGTGGGCGATCGATCTGGAACGACC
S H C E D Y A A A I A M A A G T R H G V G I D L E R P 167
1783 AATCACACCCGCGCGCGCGCGCTGTCTGAGCATCGCAATCGATGCGGACGAAGCGCTCGTCTGGCAAAGGCGGCGA
I T P A A R A A L L S I A I D A D E A A R L A K A A D 194
▼Tn5
1864 CGCGCAGTGGCGCAAGACCTGCTGCTGACCGCACTATTTTCGGCCAAGGAAGCCTGTTCAAAGCCGCTACAGCGCGT
A Q W P Q D L L L T A L F S A K E S L F K A A Y S A V 221
1945 CGGACGCTACTTCGACTTCAGCGCGSCACCGCTGTCCGCGTTCGACCTGCGCAAGCAATGCGCTGCATCTGCCCTTCGCG
G R Y F D F S A A R L C G I D L A R Q C L H L R L T E 248
2026 GACACTCTGCGCGCAATTCGTGGCGCGGCAAGTGTGCGAGGTGCGCTTCGCGGCGCTACCAACCGGACCTGGTCTCAGCCA
T L C A Q F V A G Q V C E V G F A R L P P D L V L T H 275
2107 CTACGCTGGTGAGCACGCGGACAGTCGAACCCGCCAACGCCAAGCGCACTCAAGACGTGGCGTGGCGCGTGGTCTGTC
Y A W * 278
2188 AAGCTCTCCCGCAGCGCACTCGGCGGTGGCATTGGGATTGCGGAACAGAGGTCTCAGCCCAAGCCCTGCTTGGCGAAG
2269 TCGATTTGGTGCCATCGACCACTGCAGACTGGCGGCATCGACATAAATCCGCACTCCGCTCTGCTCGAACACCGCATCG
2350 TCCGCGGTGCTCTGTCGCCAGATCGGTGACATGGCCCCAACCGGAACAGCCTGTGCGTACCAACCCGAAAGTAGACCC

FIGURE 9

**FIGURE 10**

**FIGURE 11**

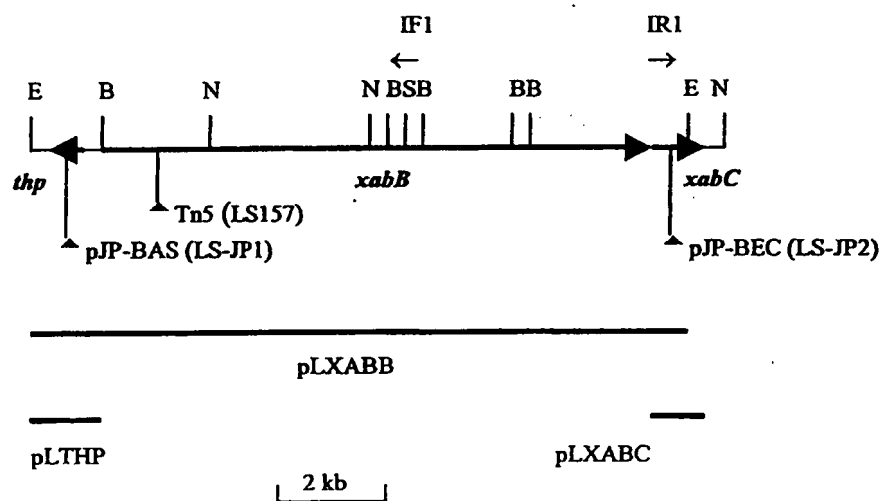


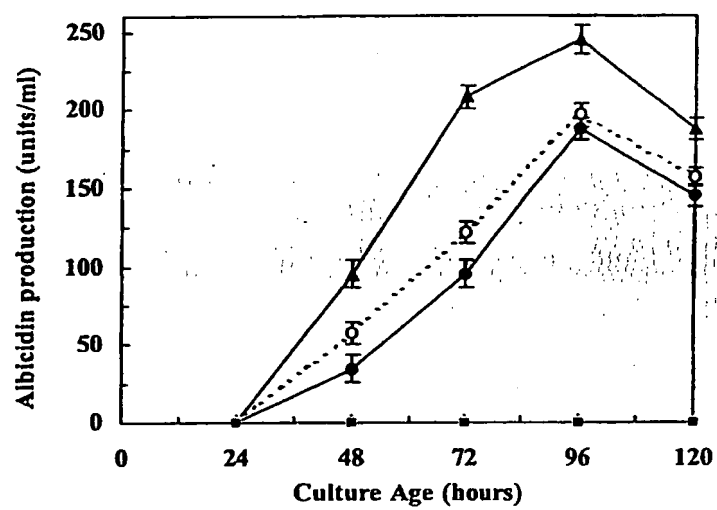
FIGURE 12

HincII +1 ClaI RBS (A3P)
mbd stop codon
 TAGCAAAAGCCGGCGCGGTACCCGTTTCATCGATAGCGAGGGCAATCATGGATTACGGCTTACCTACATCTGCATTTACCTTCGATCTCTTTACACCAAGGTTAAC 105
 V N 20 → M D S A L P T S A P T F D L P Y T T
 GCGTACTATCGCACTGCGCGAGTCAAGCGCGGATCGAACTGGGGCTATTGGATGTGGTGGGACAGGGGCGCAACTCCGCGAGCATGCGCGAGGCTGCGAGGCGTGGCGGC
 GCGGC 225
 A Y Y R T A A V K A A I E L G L F D V V G Q Q C R T P A A I A E A C Q A S P R
 G 60
 ClaI
 ATTGGCATCCTTTGCTATTACCTAGTATCGATGGGTTTCTACGGCGCAAGGGTGGCTGTCTACATAGATCGCAACATGGCCATGTACCTGGATCGTAGTTGCGCGGCTACC
 TGGGT 345
 I R I L C Y Y L V S I G F L R R N G G L P Y I D R N M A N Y L D R S S P C Y L
 G 100
 GCGCATCAAGTTCCTGCTCTCGGCTTACATCATGAGCGCTTTCACGATCTGACCGCGCTACTCAGGACCGCAAGATCAACCTGGCGAGGAGCGCGTGGTGGCAAGGATC
 ACCCG 465
 G S I K P L L S P Y I M S A P T D L T A V V R T G K I N L A Q D G V V A P D H
 P 140
 CAGTGGGTGGANTTTGCAAGGCGATGGCAAGCATGATGGGCGTGGCTGGGCTGGATCGGCAATATGGTGTGCTTCCCGCTGATCGGCGGATTCGCTGCTGCTGCGAGTGGCAG
 CCGGC 585
 Q W V E F A R A M A P M M A L P S A L I A N M V S L P A D R P I R V L D V A
 A G 180
 Motif
 I
 CAGCGCTGTTCGGCATCGCTTCGGCGAGCGCTTCGGCGAGCGCTGAGTGAGCTTCTGCACTGGGACACGCTGTAGACGTAGCAGCGGAAACGGCGCGCGCGCAAGTGG
 CCGAG 705
 H G L P G I A F A Q R P R Q A E V S P L D W D N V L D V A R E N A Q A A K V
 A E 220
 (IR)
 HincII
 CGAGCGGCTTTCGCGCGGCAAGCGATTGCACTGCAATTAGCGGAGCGCTTACGAGTGTATCTTTGACCAACTTCTTGCACCATTTGATGAGGTGATGGGAGCGCATCT
 TGGCT 825
 R A R P L P G N A F D L D Y G S G Y D V I L L T N P L H R F D E V D G E R I
 L A 260
 EcoRI Motif II
 AAGACGGCGGATGCGCTGAAAGAGGAGCGGATGGTATCACTTTGCAATTATCGCGAGAGCGGTTCTTACCGCGCGTGGCGCGCACCTTACGATGATGATGCTGGCA
 CCAC 945
 K T R D A L N D D G M V I P E P I A D E E R S S P P L A A T P S N M M L G
 T T 300
 Motif III
 ClaI
 CCGCGGGCGAGTCTACACCTATAGCGATCGGAAGGATGTTTGGCATCGCGGCTTGGCCACGTGGAACATAAAATCGATACCGCGCGGCTTCTGAAAGTGGTGGTTTCC
 CCAAG 1065
 GGTTC
 P A G R S Y T Y S D L E R M F R H A G P G H V E L K S I P P A L L K V V V S R
 V 240
 AGGGGCGCAATGATGCAATGGGCGACATCCCTTGGGCGAAACCGAGCGCATCTGCTGCAACCGAGCTGGACCTGGATGCACTCAAGCCATGTGGGCAACAGGATGCAAGC
 CTGC 1185
 TCCCGGGTATTACT (A3R)
 R A P -
 343
 TCGGTATACGATGATGAGATCGGCTGGACTATCTGGTCTCTGCAATGTGGTGGACTGGGTTGCCACCGCCCTATGGGGTATTGCAATGGGGGCGATCGCTCACCTGGC
 CGAGG 1305
 NotI
 CTACCGCGAGATGGCGGCTCCATGTGGTGGCGCGCGGCAAGTGGTGGCTAGACATCAATGCCAACACATCGCGAGCATCTCCAGTGGCGCAAGTACAGTGCATCC
 GCGC 1425
 CCGTCGCATAGGGGCGCTTGACCGAGGTATGGCAGATGGCATCTATGACGAGGTGACCGCAGATCTGGTGTGGGCGCTGACCATGG
 1515
 HcoI

FIGURE 13

Xal-XabC	174	VLDVAAGHG	236	SGYDVILL	267	ALNDDGMVIT
Sgl-TcmO	173	FVDLGGARG	234	PRADVFIV	263	ALTPGGAVLV
Sgl-TcmN	331	IADLGGGDG	393	TGYDAYLF	423	IGDDDARLLI
Smy-MdmC	64	VLEIGTFTG	135	GAFDIVFV	159	LVRPGGLVAI
Mxa-SafC	63	TLEVGCVFTG	134	GTFDLAFI	158	LVRPGGLIIL
Ser-EryG	85	VLDVGFGLG	149	ETFDRVTS	178	VLKPGGVLAI
Spe-DauK	183	VLDVGGGKG	254	RKADAILL	273	ALEPGGRILI
Sal-DmpM	208	VVDIGGADG	269	GGGDLYVL	298	AMPAHARLLV
Shy-RapM	106	VLEVGCGMG	155	VQGDAEEL	194	ALRRGGALSH
Sav-AveD	71	VLDVGCGSG	124	GSFDAAWA	151	LVRPGGRLAV
		Motif I		Motif II		Motif III

FIGURE 14

**FIGURE 15**

SEQUENCE LISTING

<110> The University of Queensland (All designated states, except U.S.)
Robert, Birch (U.S. only)

<120> Polynucleotides and polypeptides associated with antibiotic
biosynthesis and uses therefor

<130> 2454928/VPA

<140> Not yet assigned

<141> 2001-09-21

<150> AU PR0277/00

<151> 2000-09-21

<150> AU PR0304/00

<151> 2000-09-22

<150> AU PR0320/00

<151> 2000-09-22

<160> 107

<170> PatentIn version 3.1

<210> 1

<211> 16511

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1230) .. (15632)

<223>

<220>

<221> 5'UTR

<222> (1) .. (1229)

<223>

<220>

<221> 3'UTR

<222> (15633) .. (16511)

<223>

<400> 1

gaattcatcg ccggccagac gccgtgcaa agggatcatgg aacagcgctt gctcccgctc 60

gctttccagc gcccgcatgc ctgccacca taaagcggtt ctttcgatat ctctcatgca 120

tacgtccgg ttcgtggtcg gcttgcccg atgcatcata gatatgcatg actcgattcg 180

cggcaccgtg ccttgatggt gggtgcaa cgaaaacaat aaccaaaggg tgggtgctcga 240

cggttttact gttagcgacac cttgtccatc gccttacgga tgggtctgac cagcgaagcg 300

aaagatgaga taaaccacat cagctgtcaa cgccgattta aatttgaccc actttccttt 360

gaatcgtcga agtaaactctg acccaccggtg ggtcttccat cgtcgggctg ctaggctgcg 420
 cagggcaaaag cccgtcgcag cccagcagcc ctgcgccggc tcacgcccga agggcaggta 480
 gccgatctcg tcgaccacca gcagcttcgg ccctagtacc gcgcgattga agtagtcctt 540
 cagccgggttc tgcgccttga ccgctgccag ttgcatcatc aggtcggccg cggatgatgaa 600
 acgtgccttg tgccccgcca tcaccgcacg ctggcacagc gccagggcga tgtgggtctt 660
 gccgacaccg ctggggccaa gcataccac gttctcggcg cgctcgacga aggtcagggtg 720
 gccgagctcg acgatctgcg ccttcgaggc gccgccggcc tggggccagt cgaactgctc 780
 cagcgtcttg atggacggca tcctggcaag tcgcgtcagc accgtgcgct tgcgctcttc 840
 acgcgcgagc tgctcgttg ccagcacctt ctccaggaag tagctggcat cctcgcacgc 900
 ggcggcctgt gcgagtgtt gccagtccga gctcaggcgt gccagcttca actgctcgca 960
 cagcgcggcg atgcgcgcac actgcaggtc catcacgcca cctccagcag ggtgtcatac 1020
 acggccagcg gatgctgcag gttttccact ggcagggcca ctggctgtcg taagggaagc 1080
 ggtgccttga gcgccggtgc ggacagtata acgacacgtt ccttgggcaa gcgcactgtc 1140
 ggcacggcct tgctgatgcc gcccatgtag ccgcgcgcct ggatctcgcg tagtagcacc 1200
 acgctggcgg ggatccatcg agggcgcgc ttg ccc aat gcg ctc atg cag ata 1253
 Leu Pro Asn Ala Leu Met Gln Ile
 1 5
 act ctt gta gcc gtc cag ttt gca ggc gta ttg tta ggc gtc acc gct 1301
 Thr Leu Val Ala Val Gln Phe Ala Gly Val Leu Leu Gly Val Thr Ala
 10 15 20
 cgc gcg gcg atc ccc aat aag gcg ggt atg aga cgc gca tgg ccg ccc 1349
 Arg Ala Ala Ile Pro Asn Lys Ala Gly Met Arg Arg Ala Trp Pro Pro
 25 30 35 40
 ttc ccg cag gcg tgc tgt cgc tct att gct tac ctc atg cag aga tcg 1397
 Phe Pro Gln Ala Cys Cys Arg Ser Ile Ala Tyr Leu Met Gln Arg Ser
 45 50 55
 cca atg tcg ccg tta cag caa acg ctg cta acc cgc ctc gcc agt gcg 1445
 Pro Met Ser Pro Leu Gln Gln Thr Leu Leu Thr Arg Leu Ala Ser Ala
 60 65 70
 gcc gcc tcc ccg aca atg atc gag ttt ccg cgt ccg gag cac gca tcg 1493
 Ala Ala Ser Arg Thr Met Ile Glu Phe Pro Arg Pro Glu His Ala Ser
 75 80 85
 cca caa tgt tgc gac gat gcc gag ctt gcg cga ctg atc gtg cag ttg 1541
 Pro Gln Cys Cys Asp Asp Ala Glu Leu Ala Arg Leu Ile Val Gln Leu
 90 95 100
 tcg gcg gga ctg caa ccg ctg gcg atg ccg ggt acc tac gtg atc att 1589
 Ser Ala Gly Leu Gln Pro Leu Ala Met Pro Gly Thr Tyr Val Ile Ile
 105 110 115 120
 gcc gcg cca cat ggt ggt ttg ttc gcg gca gcc ctg ctt gcc tgt ttg 1637
 Ala Ala Pro His Gly Gly Leu Phe Ala Ala Ala Leu Leu Ala Cys Leu
 125 130 135

cat gcc aac ctg gtg gcg gtg ccg ttt cca ctg gat gtt gct cag cca His Ala Asn Leu Val Ala Val Pro Phe Pro Leu Asp Val Ala Gln Pro 140 145 150	1685
aat gag cgg gaa cag gcc agg ctg gag acg atc cac gca caa ttg atg Asn Glu Arg Glu Gln Ala Arg Leu Glu Thr Ile His Ala Gln Leu Met 155 160 165	1733
gag cat ggc aat gta gcg gtt ctg ctt gac gat gtc gcc gat cgc agt Glu His Gly Asn Val Ala Val Leu Leu Asp Asp Val Ala Asp Arg Ser 170 175 180	1781
gcc ttc gcg cgc atg gcg cat gct gcg ggc acc ttc ctg gcg acc ttc Ala Phe Ala Arg Met Ala His Ala Ala Gly Thr Phe Leu Ala Thr Phe 185 190 195 200	1829
gcc gat cta aag cgc gaa tcg acc agc gcc tcc ttg tgc ccg gcg tcg Ala Asp Leu Lys Arg Glu Ser Thr Ser Ala Ser Leu Cys Pro Ala Ser 205 210 215	1877
cct tcg gac gcc gcc ttg ctg ttg ttt acc tct ggt tcc tcg ggt gag Pro Ser Asp Ala Ala Leu Leu Leu Phe Thr Ser Gly Ser Ser Gly Glu 220 225 230	1925
tcc aag ggc atc ctg ctt agc cac cgc aac ctg cat cat cag atc cag Ser Lys Gly Ile Leu Leu Ser His Arg Asn Leu His His Gln Ile Gln 235 240 245	1973
gct ggc atc cgg cag tgg agc ttg gac gag cat agc cat gtg gtg acc Ala Gly Ile Arg Gln Trp Ser Leu Asp Glu His Ser His Val Val Thr 250 255 260	2021
tgg ctt tct ccc gcg cac aac ttc ggc ctg cat ttc ggc ttg ctg gca Trp Leu Ser Pro Ala His Asn Phe Gly Leu His Phe Gly Leu Leu Ala 265 270 275 280	2069
ccc tgg ttc agt ggc gcg acg gtc agt ttc atc cat ccg cac agt tat Pro Trp Phe Ser Gly Ala Thr Val Ser Phe Ile His Pro His Ser Tyr 285 290 295	2117
atg aaa cga ccc ggc ttc tgg ctg gag acg gtt gcg gct aga gac gcc Met Lys Arg Pro Gly Phe Trp Leu Glu Thr Val Ala Ala Arg Asp Ala 300 305 310	2165
acg cac atg gcc gcg ccg aac ttc gcg ttc gac tac tgc tgc gac tgg Thr His Met Ala Ala Pro Asn Phe Ala Phe Asp Tyr Cys Cys Asp Trp 315 320 325	2213
gtg atg gtc gag cag ctt ccg ccg tct gcg ttg tct acg ctt acg cat Val Met Val Glu Gln Leu Pro Pro Ser Ala Leu Ser Thr Leu Thr His 330 335 340	2261
atc gtg tgt ggc ggc gag ccg gtg cgc gcc tcg acc atg cag cgc ttc Ile Val Cys Gly Gly Glu Pro Val Arg Ala Ser Thr Met Gln Arg Phe 345 350 355 360	2309
ttc gag aaa ttc gcc gga ctc ggt gcg cgt acg cag act ttc atg ccg Phe Glu Lys Phe Ala Gly Leu Gly Ala Arg Thr Gln Thr Phe Met Pro 365 370 375	2357
cac ttc ggc ttg tct gaa acc ggt gcg ctg agt acc ttg gac gag gcg His Phe Gly Leu Ser Glu Thr Gly Ala Leu Ser Thr Leu Asp Glu Ala	2405

380	385	390	
ccc caa cag cgc gtc ttg gaa cta gat gcc gac gcc ttg aac aaa cgc			2453
Pro Gln Gln Arg Val Leu Glu Leu Asp Ala Asp Ala Leu Asn Lys Arg			
395	400	405	
aag cgc gtg gcg gca ggg gcg agc cag gcg cgt gtg aca gtg ctc aat			2501
Lys Arg Val Ala Ala Gly Ala Ser Gln Ala Arg Val Thr Val Leu Asn			
410	415	420	
tgc ggc gcc gtc gac caa gat gtg gag ttg cgt atc gtc tgt cct gaa			2549
Cys Gly Ala Val Asp Gln Asp Val Glu Leu Arg Ile Val Cys Pro Glu			
425	430	435	440
ggc gag acg ttg tgc aga cca gat gag atc ggc gaa ata tgg gta aag			2597
Gly Glu Thr Leu Cys Arg Pro Asp Glu Ile Gly Glu Ile Trp Val Lys			
445	450	455	
tcg cct gcg atc gcc cgt ggc tac ctg ttt gcg aag ccc gcc gat cag			2645
Ser Pro Ala Ile Ala Arg Gly Tyr Leu Phe Ala Lys Pro Ala Asp Gln			
460	465	470	
cga cag ttc aac tgc agc atc cgt cat acc gac gat agc ggt tac ttt			2693
Arg Gln Phe Asn Cys Ser Ile Arg His Thr Asp Asp Ser Gly Tyr Phe			
475	480	485	
cgt acc ggc gac ctg ggt ttc att gcc gat ggc tgt ctg tat gtc acc			2741
Arg Thr Gly Asp Leu Gly Phe Ile Ala Asp Gly Cys Leu Tyr Val Thr			
490	495	500	
gga agg gta aag gag gtg ctg atc ata cgc ggt aag aat cat tac ccc			2789
Gly Arg Val Lys Glu Val Leu Ile Ile Arg Gly Lys Asn His Tyr Pro			
505	510	515	520
gca cat atc gaa gcc tcg atc gcc gct acc gca tcg cct ggc gcg ctg			2837
Ala His Ile Glu Ala Ser Ile Ala Ala Thr Ala Ser Pro Gly Ala Leu			
525	530	535	
atg ccg gtg gtg ttc agc atc gag cgg cag gac gag gag cgc gta gct			2885
Met Pro Val Phe Ser Ile Glu Arg Gln Asp Glu Glu Arg Val Ala			
540	545	550	
gcg gtg atc gcc gtc aat cac ccg tgg acg ccg gca gca tgc gcc gcg			2933
Ala Val Ile Ala Val Asn His Pro Trp Thr Pro Ala Ala Cys Ala Ala			
555	560	565	
cag gca cac aag atc cgg caa cag gta gcc gac cag cat gga gtc gcc			2981
Gln Ala His Lys Ile Arg Gln Gln Val Ala Asp Gln His Gly Val Ala			
570	575	580	
ctg gcg gag cta gcc ttt gcc gaa cac cgg cac gtg ttc ggc acc tat			3029
Leu Ala Glu Leu Ala Phe Ala Glu His Arg His Val Phe Gly Thr Tyr			
585	590	595	600
ccg ggc aaa ctg aag cgg cgc cta gtc aag gaa gcc tat gtc aac ggc			3077
Pro Gly Lys Leu Lys Arg Arg Leu Val Lys Glu Ala Tyr Val Asn Gly			
605	610	615	
cag ctg ccg ttg tta tgg cat gag ggt aag aac cgg gac gta cca gcg			3125
Gln Leu Pro Leu Leu Trp His Glu Gly Lys Asn Arg Asp Val Pro Ala			
620	625	630	
gcc gcc gcg gac gat cgg cag gcg caa cac gtg gcg gac ctg tgt cgg			3173

Ala Ala Ala Asp Asp Arg Gln Ala Gln His Val Ala Asp Leu Cys Arg	
635 640 645	
aag gtc ttt ttg ccg gtg ttg ggt gtc gcg ccg ccg cat gcc caa tgg	3221
Lys Val Phe Leu Pro Val Leu Gly Val Ala Pro Pro His Ala Gln Trp	
650 655 660	
ccg ctg tgc gaa ctg gcg ctg gat tgc ctc caa tgc gtg cgt ctt gcc	3269
Pro Leu Cys Glu Leu Ala Leu Asp Ser Leu Gln Cys Val Arg Leu Ala	
665 670 675 680	
ggt gcc atc gaa gag tgc tac ggc gtg cct ttc gaa ccc acg ttg cta	3317
Gly Ala Ile Glu Glu Cys Tyr Gly Val Pro Phe Glu Pro Thr Leu Leu	
685 690 695	
ttc aag ctt gag acg gtc ggg gca atc gcc gaa tat gtc ctg gcg cac	3365
Phe Lys Leu Glu Thr Val Gly Ala Ile Ala Glu Tyr Val Leu Ala His	
700 705 710	
gga cgt cag gcg ccc acg ccg acg cgt gcg ccg gtg gca agc aca aca	3413
Gly Arg Gln Ala Pro Thr Pro Thr Arg Ala Pro Val Ala Ser Thr Thr	
715 720 725	
tgc tca gag gaa ccg atc gcc att gtg gcg atg cac tgt gag gtg ccc	3461
Cys Ser Glu Glu Pro Ile Ala Ile Val Ala Met His Cys Glu Val Pro	
730 735 740	
gga_cg ggc gag aac act gaa gca ttg tgg tgc ttc ctg cgg agc gac	3509
Gly Ala Gly Glu Asn Thr Glu Ala Leu Trp Ser Phe Leu Arg Ser Asp	
745 750 755 760	
gtc aac gcg atc ccg ccg atc gaa tca acg cgc ccg gac tta tgg gca	3557
Val Asn Ala Ile Arg Pro Ile Glu Ser Thr Arg Pro Asp Leu Trp Ala	
765 770 775	
gcg atg cgc gcc cat ccc ggc ctc gcg ggc gaa cag ctg ccg cgc tat	3605
Ala Met Arg Ala Tyr Pro Gly Leu Ala Gly Glu Gln Leu Pro Arg Tyr	
780 785 790	
gcg ggt ttc ctc gac gac gtt gat gct ttc gat gct gcg ttt ttc ggt	3653
Ala Gly Phe Leu Asp Asp Val Asp Ala Phe Asp Ala Ala Phe Phe Gly	
795 800 805	
atc tgc cgt cgc gag gcc gaa tgc atg gac ccg cag cag cgc aaa gtg	3701
Ile Ser Arg Arg Glu Ala Glu Cys Met Asp Pro Gln Gln Arg Lys Val	
810 815 820	
ctg gag atg gtg tgg aag ctg atc gag caa gcc ggt cac gat ccg ctg	3749
Leu Glu Met Val Trp Lys Leu Ile Glu Gln Ala Gly His Asp Pro Leu	
825 830 835 840	
tcc tgg ggc ggc cag ccg gtc ggc ctg ttc gtg ggt gcg cat acg tcc	3797
Ser Trp Gly Gly Gln Pro Val Gly Leu Phe Val Gly Ala His Thr Ser	
845 850 855	
gac tat ggc gag ctg ctg gcg agc cag ccg caa ctg atg gcc caa tgt	3845
Asp Tyr Gly Glu Leu Leu Ala Ser Gln Pro Gln Leu Met Ala Gln Cys	
860 865 870	
ggc gct tac atc gat tgc ggt tgc cat ttg acc atg att ccg aac ccg	3893
Gly Ala Tyr Ile Asp Ser Gly Ser His Leu Thr Met Ile Pro Asn Arg	
875 880 885	

gct tgc cgc tgg ttc aat ttc acc ggc ccc agc gaa gta atc aac agc Ala Ser Arg Trp Phe Asn Phe Thr Gly Pro Ser Glu Val Ile Asn Ser 890 895 900	3941
gct tgc tcc agc tgc ctg gtg gcg ctg cat cgg gcg gtt caa tgc ctg Ala Cys Ser Ser Ser Leu Val Ala Leu His Arg Ala Val Gln Ser Leu 905 910 915 920	3989
cgc caa ggc gaa agc agt gtc gcc ctg gta ctc ggc gtg aac ctt atc Arg Gln Gly Glu Ser Ser Val Ala Leu Val Leu Gly Val Asn Leu Ile 925 930 935	4037
ctg gct ccc aag gtg ctg tta gcc agt gca agc gcg ggc atg ctt tgc Leu Ala Pro Lys Val Leu Leu Ala Ser Ala Ser Ala Gly Met Leu Ser 940 945 950	4085
ccc gat ggc cgc tgc aag acg ctt gac gcc gcc gcc gat ggc ttc gtg Pro Asp Gly Arg Cys Lys Thr Leu Asp Ala Ala Ala Asp Gly Phe Val 955 960 965	4133
cgt tgc gaa ggg atc gca ggg gtg ata ttg aag cca ctg gcg cag gcg Arg Ser Glu Gly Ile Ala Gly Val Ile Leu Lys Pro Leu Ala Gln Ala 970 975 980	4181
ctg gcc gat ggt gac agg gtc tac ggt cta gtc cgc ggc gtg gcg gtc Leu Ala Asp Gly Asp Arg Val Tyr Gly Leu Val Arg Gly Val Ala Val 985 990 995 1000	4229
aac cat ggc ggc cgt tcc aat tcc ttg cgt gct ccc aac gtc aac Asn His Gly Gly Arg Ser Asn Ser Leu Arg Ala Pro Asn Val Asn 1005 1010 1015	4274
gcg cag cgg caa ctg ctg atc cgg act tac cag gaa gcc ggt gtc Ala Gln Arg Gln Leu Leu Ile Arg Thr Tyr Gln Glu Ala Gly Val 1020 1025 1030	4319
gag ccg gcc agc gtc ggt tat gtt gaa cta cac ggc act ggt acc Glu Pro Ala Ser Val Gly Tyr Val Glu Leu His Gly Thr Gly Thr 1035 1040 1045	4364
agc ctg ggt gat ccg atc gaa atc cag gcg ctg aag gaa gct ttc Ser Leu Gly Asp Pro Ile Glu Ile Gln Ala Leu Lys Glu Ala Phe 1050 1055 1060	4409
att gcg ttg ggg gca cag gcc gcc ccg tca aac tgc ggc atc ggt Ile Ala Leu Gly Ala Gln Ala Ala Pro Ser Asn Cys Gly Ile Gly 1065 1070 1075	4454
tgc gtg aag tcc gcg ctg ggc cat cta gaa gcc gct gca ggc ctg Ser Val Lys Ser Ala Leu Gly His Leu Glu Ala Ala Ala Gly Leu 1080 1085 1090	4499
acc ggc ctg atc aag gtg ctg ctg atg ctc aag cac ggc gag cag Thr Gly Leu Ile Lys Val Leu Leu Met Leu Lys His Gly Glu Gln 1095 1100 1105	4544
gcc ggc acg cgc cat ttc agc acg ctc aat ccg ctg atc gat ttg Ala Gly Thr Arg His Phe Ser Thr Leu Asn Pro Leu Ile Asp Leu 1110 1115 1120	4589
cga ggt acg tca ttc gaa gtg gtg gcg cag cat cgc gca tgg ccg Arg Gly Thr Ser Phe Glu Val Val Ala Gln His Arg Ala Trp Pro 1125 1130 1135	4634

tcg cag gtc ggc att	cac ggc aca ctc ttg	ccg cgt cgc gcg ggt	4679
Ser Gln Val Gly Ile	His Gly Thr Leu Leu	Pro Arg Arg Ala Gly	
1140	1145	1150	
atc agc tca ttc ggc	ttc ggc ggc gcc aat	gcg cat gcg atc gtg	4724
Ile Ser Ser Phe Gly	Phe Gly Gly Ala Asn	Ala His Ala Ile Val	
1155	1160	1165	
gaa gag cat gtc att	gcc acg ccc ccc tcg	acg agc tcc gct ggc	4769
Glu Glu His Val Ile	Ala Thr Pro Pro Ser	Thr Ser Ser Ala Gly	
1170	1175	1180	
ggc ccg gta ggt atc	gtg ttg tca gcc ggt	agt gaa gct gtc ttg	4814
Gly Pro Val Gly Ile	Val Leu Ser Ala Gly	Ser Glu Ala Val Leu	
1185	1190	1195	
cgg caa caa gtg ctg	gcc ttg tca gcc tgg	cta agg cag caa tcg	4859
Arg Gln Gln Val Leu	Ala Leu Ser Ala Trp	Leu Arg Gln Gln Ser	
1200	1205	1210	
ccg aca ccc gcg caa	atg atc gat gtc gcc	tac acc tta cag gta	4904
Pro Thr Pro Ala Gln	Met Ile Asp Val Ala	Tyr Thr Leu Gln Val	
1215	1220	1225	
gga cgc gca gcc ctg	tcg cac agg ttg gct	ttt agc gcg acg gac	4949
Gly Arg Ala Ala Leu	Ser His Arg Leu Ala	Phe Ser Ala Thr Asp	
1230	1235	1240	
gcc gag cag gca ttg	gcg agg ctt gag ggt	cgt ctg gcg ggc gtg	4994
Ala Glu Gln Ala Leu	Ala Arg Leu Glu Gly	Arg Leu Ala Gly Val	
1245	1250	1255	
atg gat gcc gag gtc	cat cac ggt gtc gtg	gat gct gcc gca acg	5039
Met Asp Ala Glu Val	His His Gly Val Val	Asp Ala Ala Ala Thr	
1260	1265	1270	
gct ccc gaa cat ggg	cgg cag acg cgc gaa	ggt ctt gcc ggt ttg	5084
Ala Pro Glu His Gly	Arg Gln Thr Arg Glu	Gly Leu Ala Gly Leu	
1275	1280	1285	
ctg cga gcc tgg act	cag ggc gtg cgc gtc	gat tgg tcg gcg ctg	5129
Leu Arg Ala Trp Thr	Gln Gly Val Arg Val	Asp Trp Ser Ala Leu	
1290	1295	1300	
tac ggc ata cag cga	ccg cag cgc gtt agc	ctg cct gtc tac ccc	5174
Tyr Gly Ile Gln Arg	Pro Gln Arg Val Ser	Leu Pro Val Tyr Pro	
1305	1310	1315	
ttc gct agg gaa cgc	tat tgg ctg ccc ggc	cag gct atg cat gcc	5219
Phe Ala Arg Glu Arg	Tyr Trp Leu Pro Gly	Gln Ala Met His Ala	
1320	1325	1330	
gct gcg gac gct cat	ccg atg ctg cag ctg	ttg cat gcc aat gcc	5264
Ala Ala Asp Ala His	Pro Met Leu Gln Leu	Leu His Ala Asn Ala	
1335	1340	1345	
aaa cta cat cgc tac	gcc ttg cgt agg tcc	ggc tgc gca agc ttt	5309
Lys Leu His Arg Tyr	Ala Leu Arg Arg Ser	Gly Cys Ala Ser Phe	
1350	1355	1360	
ctt gtt gat cat tgc	gtg gat ggt cga cag	gta cta ccg gca gcc	5354
Leu Val Asp His Cys	Val Asp Gly Arg Gln	Val Leu Pro Ala Ala	

1365	1370	1375	
gtg caa ctg gaa ttg	gtg cgc gcc gtg gcg	cag cgg gtc atg gcg	5399
Val Gln Leu Glu Leu	Val Arg Ala Val Ala	Gln Arg Val Met Ala	
1380	1385	1390	
cag gat gag ggt tgt	atc gaa ctg gcg cag	gtc gcc ttt ttg cat	5444
Gln Asp Glu Gly Cys	Ile Glu Leu Ala Gln	Val Ala Phe Leu His	
1395	1400	1405	
ccc ctc atg atg gag	gag act gag ctg gag	gtc gaa atc gaa ctg	5489
Pro Leu Met Met Glu	Glu Thr Glu Leu Glu	Val Glu Ile Glu Leu	
1410	1415	1420	
tcg aag agc gat caa	gat gag ttc gat ttc	caa ctt cac gat gct	5534
Ser Lys Ser Asp Gln	Asp Glu Phe Asp Phe	Gln Leu His Asp Ala	
1425	1430	1435	
cac cgc caa cag gtc	ttt agc cag ggg cac	gta cgt cgc cgg gtc	5579
His Arg Gln Gln Val	Phe Ser Gln Gly His	Val Arg Arg Arg Val	
1440	1445	1450	
tat acg gcg aca ccg	cgc ttg gat tta gcc	cag ctg caa aag ctt	5624
Tyr Thr Ala Thr Pro	Arg Leu Asp Leu Ala	Gln Leu Gln Lys Leu	
1455	1460	1465	
tgt gcc gag cgc gtg	ttg tcc ggc gaa gac	tgt tat gcg cac ttc	5669
Cys Ala Glu Arg Val	Leu Ser Gly Glu Asp	Cys Tyr Ala His Phe	
1470	1475	1480	
acc gcc tgc gga ttg	cag ctc ggc gac cgg	ctc aaa tcc gtg caa	5714
Thr Ala Cys Gly Leu	Gln Leu Gly Asp Arg	Leu Lys Ser Val Gln	
1485	1490	1495	
tcg atc ggc tgc gga	cgc aat ggc gag ggc	gag ccg atc gca ttg	5759
Ser Ile Gly Cys Gly	Arg Asn Gly Glu Gly	Gln Pro Ile Ala Leu	
1500	1505	1510	
ggt gtc ctg cgc ctg	cca cca tca agc gtt	gaa gac agc cat gtg	5804
Gly Val Leu Arg Leu	Pro Pro Ser Ser Val	Glu Asp Ser His Val	
1515	1520	1525	
ctg cct cct agc ctg	ctt gat ggt gcc ttg	cag tgt agc ctt ggc	5849
Leu Pro Pro Ser Leu	Leu Asp Gly Ala Leu	Gln Cys Ser Leu Gly	
1530	1535	1540	
ttg cag cgt gat gtc	gag cac atc gcc atg	cca tac acg ctg gag	5894
Leu Gln Arg Asp Val	Glu His Ile Ala Met	Pro Tyr Thr Leu Glu	
1545	1550	1555	
cgg atg acg gtg cat	gcg ccg att cct ccc	gag gcc tgg gtg ctg	5939
Arg Met Thr Val His	Ala Pro Ile Pro Pro	Glu Ala Trp Val Leu	
1560	1565	1570	
ctg cgt cac ggc cat	gca gcc aga cag tcc	ctg gac atc gat ctc	5984
Leu Arg His Gly His	Ala Ala Arg Gln Ser	Leu Asp Ile Asp Leu	
1575	1580	1585	
ctg gat tcc gaa ggt	agg gtc tgc gtc agc	ctc ggc aat tac acc	6029
Leu Asp Ser Glu Gly	Arg Val Cys Val Ser	Leu Gly Asn Tyr Thr	
1590	1595	1600	
ggc cgt gca ccg aaa	gcc gtt tcc gcc gtc	agg gcg ctt gtc ttg	6074

Gly Arg Ala Pro Lys	Ala Val Ser Ala Val	Arg Ala Leu Val Leu	
1605	1610	1615	
gca ccg gtc tgg caa	gcg ttg acc gaa acg	gcg ccg gca tgg ccc	6119
Ala Pro Val Trp Gln	Ala Leu Thr Glu Thr	Ala Pro Ala Trp Pro	
1620	1625	1630	
gat ccg gcc gaa cgc	atc gtt acg gta gga	gac gat gca tgg cgt	6164
Asp Pro Ala Glu Arg	Ile Val Thr Val Gly	Asp Asp Ala Trp Arg	
1635	1640	1645	
agt cac ttc ggt ttc	gac gag ccg gcc ttg	tcc ctg gag gac agc	6209
Ser His Phe Gly Phe	Asp Glu Pro Ala Leu	Ser Leu Glu Asp Ser	
1650	1655	1660	
gtc gaa gtc atc gcg	acg cga ctg ggc cag	agc ggc aag ttc gat	6254
Val Glu Val Ile Ala	Thr Arg Leu Gly Gln	Ser Gly Lys Phe Asp	
1665	1670	1675	
cat cta gtc tgg atc	gtg ccg ata gcc gag	agt gaa acc gat att	6299
His Leu Val Trp Ile	Val Pro Ile Ala Glu	Ser Glu Thr Asp Ile	
1680	1685	1690	
gca gcg caa ggt tca	gcg gcg atc gcc ggt	ttc cgg ttg gtc aag	6344
Ala Ala Gln Gly Ser	Ala Ala Ile Ala Gly	Phe Arg Leu Val Lys	
1695	1700	1705	
gcg ttg ctt gcg ttg	ggc tat gcg cat cgc	ccg ctg ggt ctc acc	6389
Ala Leu Leu Ala Leu	Gly Tyr Ala His Arg	Pro Leu Gly Leu Thr	
1710	1715	1720	
gtg ctg act cgc caa	gcc ctt acg cgg cag	ccg tcg cac gcg gca	6434
Val Leu Thr Arg Gln	Ala Leu Thr Arg Gln	Pro Ser His Ala Ala	
1725	1730	1735	
gtg cac ggg ctg atc	ggg acg ccg gcc aag	gaa tac tgc aac ttg	6479
Val His Gly Leu Ile	Gly Thr Leu Ala Lys	Glu Tyr Cys Asn Trp	
1740	1745	1750	
aaa atc cgt ctg ctc	gac ctg ccg agc gta	aaa tct tgg ccg caa	6524
Lys Ile Arg Leu Leu	Asp Leu Pro Ser Val	Lys Ser Trp Pro Gln	
1755	1760	1765	
tgg gag caa ttg cgg	tcg ttg cct tgg cat	gcg cag ggc gaa gcc	6569
Trp Glu Gln Leu Arg	Ser Leu Pro Trp His	Ala Gln Gly Glu Ala	
1770	1775	1780	
ctg atc ggc cgt ggg	act tgt tgg tat cgg	cgg cag ttg tgt gaa	6614
Leu Ile Gly Arg Gly	Thr Cys Trp Tyr Arg	Arg Gln Leu Cys Glu	
1785	1790	1795	
gtg ctg ccg ctg ccg	tcg ttg gaa ccg ccg	ccg tac cgc gta ggc	6659
Val Leu Pro Leu Pro	Ser Leu Glu Pro Pro	Pro Tyr Arg Val Gly	
1800	1805	1810	
ggt gtc tac gtc gtg	atc ggc ggc gct ggc	ggc ttg ggt gaa gta	6704
Gly Val Tyr Val Val	Ile Gly Gly Ala Gly	Gly Leu Gly Glu Val	
1815	1820	1825	
ttg agc gaa cac ttg	atc cgc acg tac gac	gcg cag ctg atc tgg	6749
Leu Ser Glu His Leu	Ile Arg Thr Tyr Asp	Ala Gln Leu Ile Trp	
1830	1835	1840	

atc ggg cgg cgc gtg	ctg gac gaa ggc att	gcg cgc aag cag acc	6794
Ile Gly Arg Arg Val	Leu Asp Glu Gly Ile	Ala Arg Lys Gln Thr	
1845	1850	1855	
cgg ctt gcg tcg ctg	ggc cgc gca ccg cat	tac atc tcc gcg gac	6839
Arg Leu Ala Ser Leu	Gly Arg Ala Pro His	Tyr Ile Ser Ala Asp	
1860	1865	1870	
gcg agt gac ccg gct	gcc ctg cag gcg gca	cat aat gag atc gtt	6884
Ala Ser Asp Pro Ala	Ala Leu Gln Ala Ala	His Asn Glu Ile Val	
1875	1880	1885	
gcg ctg cat ggc cag	ccc cat ggg ctc atc	cta agc aac atc gtg	6929
Ala Leu His Gly Gln	Pro His Gly Leu Ile	Leu Ser Asn Ile Val	
1890	1895	1900	
ctg aag gat gcc agt	ctg gct cgt atg gag	gaa gcc gat ttc cgt	6974
Leu Lys Asp Ala Ser	Leu Ala Arg Met Glu	Glu Ala Asp Phe Arg	
1905	1910	1915	
gac gtg ctg gcc gcg	aaa ctc gac gtc agc	gtg tgt gcg gca cag	7019
Asp Val Leu Ala Ala	Lys Leu Asp Val Ser	Val Cys Ala Ala Gln	
1920	1925	1930	
gtg ttc ggc acg gcc	ccc ctt gat ttc gtg	ctg ttt ttt tct tcc	7064
Val Phe Gly Thr Ala	Pro Leu Asp Phe Val	Leu Phe Phe Ser Ser	
1935	1940	1945	
atc cag agc act acc	aag gcg gcc ggg caa	ggg aac tac gcc gcc	7109
Ile Gln Ser Thr Thr	Lys Ala Ala Gly Gln	Gly Asn Tyr Ala Ala	
1950	1955	1960	
ggc tgc tgc tat gtc	gac gct ttc ggc gag	cta tgg gcg cgc cgg	7154
Gly Cys Cys Tyr Val	Asp Ala Phe Gly Glu	Leu Trp Ala Arg Arg	
1965	1970	1975	
ggg ttg agg gta aag	acc atc aac tgg ggc	tac tgg ggc agc gtg	7199
Gly Leu Arg Val Lys	Thr Ile Asn Trp Gly	Tyr Trp Gly Ser Val	
1980	1985	1990	
ggc gtc gta gcg gcc	gag gac tat cgc cgg	cgc atg gcg caa aaa	7244
Gly Val Val Ala Gly	Glu Asp Tyr Arg Arg	Arg Met Ala Gln Lys	
1995	2000	2005	
cac atg gct tcg att	gag ggt gcc gaa gcg	atg cag gtg ttg tcg	7289
His Met Ala Ser Ile	Glu Gly Ala Glu Ala	Met Gln Val Leu Ser	
2010	2015	2020	
cag ttg ttg tgt gcg	ccg ttg caa cgg ctt	gcc tac gtc aag atc	7334
Gln Leu Leu Cys Ala	Pro Leu Gln Arg Leu	Ala Tyr Val Lys Ile	
2025	2030	2035	
gac gat gct aac gca	atg cgc gct ctg ggc	gta gta gag gac gag	7379
Asp Asp Ala Asn Ala	Met Arg Ala Leu Gly	Val Val Glu Asp Glu	
2040	2045	2050	
agc gtg caa atc cct	gtg cac gca ccg gcc	gag cct ccc aga ggg	7424
Ser Val Gln Ile Pro	Val His Ala Pro Ala	Glu Pro Pro Arg Gly	
2055	2060	2065	
cag cct ggt ccc gtg	gtc gag ttg tcg gtg	aat ctg gat gcc cgg	7469
Gln Pro Gly Pro Val	Val Glu Leu Ser Val	Asn Leu Asp Ala Arg	
2070	2075	2080	

cgc gaa cgg gaa act	ttg ctg gcg gcc tgg	ctg ctt gag ttg atc	7514
Arg Glu Arg Glu Thr	Leu Leu Ala Ala Trp	Leu Leu Glu Leu Ile	
2085	2090	2095	
gag caa ctc ggt ggt	ttt ccg ccg gca agt	ttc gac atc gct acg	7559
Glu Gln Leu Gly Gly	Phe Pro Pro Ala Ser	Phe Asp Ile Ala Thr	
2100	2105	2110	
ctt gcg caa cgc ctg	cac atc gta ccc gcc	tat cga agc tgg ctg	7604
Leu Ala Gln Arg Leu	His Ile Val Pro Ala	Tyr Arg Ser Trp Leu	
2115	2120	2125	
gaa cac agc gtg cgg	atg ctc ggc gtg tat	ggg tac ctc aga gcg	7649
Glu His Ser Val Arg	Met Leu Gly Val Tyr	Gly Tyr Leu Arg Ala	
2130	2135	2140	
acg ggg gaa agc cga	ttc gag ctg gcc gac	aag ccg ccc gat gat	7694
Thr Gly Glu Ser Arg	Phe Glu Leu Ala Asp	Lys Pro Pro Asp Asp	
2145	2150	2155	
gcc agg ggt gcc tgg	aac gcg cat gtg cac	gag gcc agc gtc gaa	7739
Ala Arg Gly Ala Trp	Asn Ala His Val His	Glu Ala Ser Val Glu	
2160	2165	2170	
gcc ggt gaa gag gca	cag ccg cgt ctg ctc	gat cgc tgc atg cgg	7784
Ala Gly Glu Glu Ala	Gln Arg Arg Leu Leu	Asp Arg Cys Met Arg	
2175	2180	2185	
gcg ttg ccg gcg gtc	ctt cga ggc gaa cgc	aag gcc acc gaa ttg	7829
Ala Leu Pro Ala Val	Leu Arg Gly Glu Arg	Lys Ala Thr Glu Leu	
2190	2195	2200	
ctg ttt ccg gaa ggt	tcg atg gcg tgg gtc	gag ggt atc tac cag	7874
Leu Phe Pro Glu Gly	Ser Met Ala Trp Val	Glu Gly Ile Tyr Gln	
2205	2210	2215	
aac aac ccg ctt gcc	gat tac ttc aac gca	caa cta gtc acg cga	7919
Asn Asn Pro Leu Ala	Asp Tyr Phe Asn Ala	Gln Leu Val Thr Arg	
2220	2225	2230	
ctg att gcc tac ttg	aga cga cga cta gag	tcg acg cct acg gcg	7964
Leu Ile Ala Tyr Leu	Arg Arg Arg Leu Glu	Ser Thr Pro Thr Ala	
2235	2240	2245	
cgc ctg aag ctg tgc	gag atc ggc gcc ggc	agc ggt ggt act act	8009
Arg Leu Lys Leu Cys	Glu Ile Gly Ala Gly	Ser Gly Gly Thr Thr	
2250	2255	2260	
gca agc gtg cta caa	cag ttg cag gca tat	ggg gag cat att gag	8054
Ala Ser Val Leu Gln	Gln Leu Gln Ala Tyr	Gly Glu His Ile Glu	
2265	2270	2275	
gaa tat ctc tat acc	gac ctg tcg cct gtc	ttc ctg cat cat gcg	8099
Glu Tyr Leu Tyr Thr	Asp Leu Ser Pro Val	Phe Leu His His Ala	
2280	2285	2290	
gaa aaa cac tat cag	cca cga gcg cct tat	ttg agg acc gcc tgt	8144
Glu Lys His Tyr Gln	Pro Arg Ala Pro Tyr	Leu Arg Thr Ala Cys	
2295	2300	2305	
ttc gac gta gcg cgc	gcg ccg acg gcg cag	gcc ctg gaa tct ggc	8189
Phe Asp Val Ala Arg	Ala Pro Thr Ala Gln	Ala Leu Glu Ser Gly	
2310	2315	2320	

ggc tac gac gtg gtg att gcc gcc aac gta ctg cat gct acg cgc Gly Tyr Asp Val Val Ile Ala Ala Asn Val Leu His Ala Thr Arg 2325 2330 2335	8234
gat atc gcc aag acc ttg cgc aat gcg aag gca ctc ctc aaa cct Asp Ile Ala Lys Thr Leu Arg Asn Ala Lys Ala Leu Leu Lys Pro 2340 2345 2350	8279
ggc ggt ctg ctc ttg ctc aac gaa gtg atc gag cgc agc ctc gtc Gly Gly Leu Leu Leu Leu Asn Glu Val Ile Glu Arg Ser Leu Val 2355 2360 2365	8324
ttg cac ctg act ttc ggt ctg ctg gag agc tgg tgg ttg ccc cag Leu His Leu Thr Phe Gly Leu Leu Glu Ser Trp Trp Leu Pro Gln 2370 2375 2380	8369
gac aag atc ttg cgc ctt gcc ggc tcg ccg ttg ctg gct tgc gcc Asp Lys Ile Leu Arg Leu Ala Gly Ser Pro Leu Leu Ala Cys Ala 2385 2390 2395	8414
acc tgg cgc agc ctg ctg gag gct gag ggt ttt gcg ggg ctg agc Thr Trp Arg Ser Leu Leu Glu Ala Glu Gly Phe Ala Gly Leu Ser 2400 2405 2410	8459
gtg cac agg gcg caa ccc gat gcc ggg cag gcc atc atc tgt gcc Val His Arg Ala Gln Pro Asp Ala Gly Gln Ala Ile Ile Cys Ala 2415 2420 2425	8504
tac agc gat ggg ata gtg cgg caa gcc agt acg atc gag gtt gcg Tyr Ser Asp Gly Ile Val Arg Gln Ala Ser Thr Ile Glu Val Ala 2430 2435 2440	8549
cgg aat gaa aaa gta acc gtt ccg tcg cag ccg gcg gaa gcc ggg Arg Asn Glu Lys Val Thr Val Pro Ser Gln Pro Ala Glu Ala Gly 2445 2450 2455	8594
gaa tcg ccg ctg gat ctg gtc aaa aaa ctg ctt gga cgc att ctg Glu Ser Pro Leu Asp Leu Val Lys Lys Leu Leu Gly Arg Ile Leu 2460 2465 2470	8639
aaa atg gat ccg gcc aca ctc gat acc agc cac ccg ctg gag tac Lys Met Asp Pro Ala Thr Leu Asp Thr Ser His Pro Leu Glu Tyr 2475 2480 2485	8684
tac ggt gtc gat tcg atc gtg gcg atc gaa ctg gct atg gca ctg Tyr Gly Val Asp Ser Ile Val Ala Ile Glu Leu Ala Met Ala Leu 2490 2495 2500	8729
cgc gag aca ttc ccg ggt ttt gaa gtc agc gag ctg ttt gaa acg Arg Glu Thr Phe Pro Gly Phe Glu Val Ser Glu Leu Phe Glu Thr 2505 2510 2515	8774
caa tcc atc gat acc ttg ttg ggc tct ctt gag cag gct cct ctc Gln Ser Ile Asp Thr Leu Leu Gly Ser Leu Glu Gln Ala Pro Leu 2520 2525 2530	8819
ctt gct acc ctc aca gct ccg ccg caa caa gac atg ctg cag cag Leu Ala Thr Leu Thr Ala Pro Pro Gln Gln Asp Met Leu Gln Gln 2535 2540 2545	8864
ctg aaa caa ctg ctg gcg cgt acg ctg aag ctg gac att acg cag Leu Lys Gln Leu Leu Ala Arg Thr Leu Lys Leu Asp Ile Thr Gln	8909

2550	2555	2560	
atc gac acg agc aag	acg ctg gag agc tat	ggg gtc gac tcc atc	8954
Ile Asp Thr Ser Lys	Thr Leu Glu Ser Tyr	Gly Val Asp Ser Ile	
2565	2570	2575	
gtc atc atc gaa tta	gcc aac gcc ttg cgt	gag cgc tat ccg agc	8999
Val Ile Ile Glu Leu	Ala Asn Ala Leu Arg	Glu Arg Tyr Pro Ser	
2580	2585	2590	
ttg gac gcg tca cag	ctg atg gaa acc tta	tcg atc gac cgg ctg	9044
Leu Asp Ala Ser Gln	Leu Met Glu Thr Leu	Ser Ile Asp Arg Leu	
2595	2600	2605	
gtt gcc caa tgg cag	gca acg gag ccc gcc	gta ccg gca gag cca	9089
Val Ala Gln Trp Gln	Ala Thr Glu Pro Ala	Val Pro Ala Glu Pro	
2610	2615	2620	
aca gcg gaa ccg ccg	gta gcc gac gaa gac	gcc gct gcc atc atc	9134
Thr Ala Glu Pro Pro	Val Ala Asp Glu Asp	Ala Ala Ala Ile Ile	
2625	2630	2635	
gga ctg gcc gcc cgc	ttt cca gcc gcg gac	acg ttg gag gag ttc	9179
Gly Leu Ala Gly Arg	Phe Pro Gly Ala Asp	Thr Leu Glu Glu Phe	
2640	2645	2650	
tgg aac aac ctg cgc	aac gcc caa agc agt	atg gga gag gtg cca	9224
Trp Asn Asn Leu Arg	Asn Gly Gln Ser Ser	Met Gly Glu Val Pro	
2655	2660	2665	
ggc gag cgc tgg gat	cac cag cac tac ttc	gac agt gaa cgc cag	9269
Gly Glu Arg Trp Asp	His Gln His Tyr Phe	Asp Ser Glu Arg Gln	
2670	2675	2680	
qca ccg gcc aag acg	tat agc cgc tgg ggt	gcg ttt ctg agg gac	9314
Ala Pro Gly Lys Thr	Tyr Ser Arg Trp Gly	Ala Phe Leu Arg Asp	
2685	2690	2695	
ata gac gcc ttc gat	gca gcc ttc ttt gaa	tgg ccc gac agc gtc	9359
Ile Asp Gly Phe Asp	Ala Ala Phe Phe Glu	Trp Pro Asp Ser Val	
2700	2705	2710	
gcg ctg gaa tcg gat	ccg caa gcg cgg ata	ttt cta gag cag gcc	9404
Ala Leu Glu Ser Asp	Pro Gln Ala Arg Ile	Phe Leu Glu Gln Ala	
2715	2720	2725	
tat gcc ggg atc gaa	gat gcc gcc tac acg	cct ggc tcg ctc agc	9449
Tyr Ala Gly Ile Glu	Asp Ala Gly Tyr Thr	Pro Gly Ser Leu Ser	
2730	2735	2740	
aag agc caa cgc gta	ggg gta ttc gta ggt	gtg atg aat ggt tac	9494
Lys Ser Gln Arg Val	Gly Val Phe Val Gly	Val Met Asn Gly Tyr	
2745	2750	2755	
tac agc gcc gga gcg	cgc ttc tgg caa atc	gcc aac cgc gtg tcg	9539
Tyr Ser Gly Gly Ala	Arg Phe Trp Gln Ile	Ala Asn Arg Val Ser	
2760	2765	2770	
tac cag ttc gat ttt	cgc ggg cca agc ctg	gcg gtg gat acc gcc	9584
Tyr Gln Phe Asp Phe	Arg Gly Pro Ser Leu	Ala Val Asp Thr Ala	
2775	2780	2785	
tgt tcg gct tcg ctc	acc gcg atc cac ctg	gcg ctg gaa agc ctg	9629
Cys Ser Ala Ser Leu	Thr Ala Ile His Leu	Ala Leu Glu Ser Leu	

2790	2795	2800	
cgc agc ggc agt tgc	gag gtc gca ctg gcc	ggc ggc gtg aat ctg	9674
Arg Ser Gly Ser Cys	Glu Val Ala Leu Ala	Gly Gly Val Asn Leu	
2805	2810	2815	
ctg gtc gat ccg cag	caa tat ctt aat ttg	gct ggc gcc gcg atg	9719
Leu Val Asp Pro Gln	Gln Tyr Leu Asn Leu	Ala Gly Ala Ala Met	
2820	2825	2830	
ctc tcc gcc ggc gcc	agc tgt cgg ccg ttc	ggc gag gcc gcg gac	9764
Leu Ser Ala Gly Ala	Ser Cys Arg Pro Phe	Gly Glu Ala Ala Asp	
2835	2840	2845	
ggc ttc gtg gcc ggc	gaa gcc tgc ggc gtg	gtg ctg ctc aag ccg	9809
Gly Phe Val Ala Gly	Glu Ala Cys Gly Val	Val Leu Leu Lys Pro	
2850	2855	2860	
ctc aag caa gcg agg	gcc gat ggc gat gtg	atc cat gcc gta atc	9854
Leu Lys Gln Ala Arg	Ala Asp Gly Asp Val	Ile His Ala Val Ile	
2865	2870	2875	
agg ggc agc atg atc	aat gcc ggt ggg cac	acc agc gcg ttc tcc	9899
Arg Gly Ser Met Ile	Asn Ala Gly Gly His	Thr Ser Ala Phe Ser	
2880	2885	2890	
tcg cct aac cct gcc	gcc cag gcc gaa gtc	gtg cgg cag gcc ttg	9944
Ser Pro Asn Pro Ala	Ala Gln Ala Glu Val	Val Arg Gln Ala Leu	
2895	2900	2905	
cag cgc gcg ggc gtg	gcg ccc gat tcg atc	agc tac atc gag gcg	9989
Gln Arg Ala Gly Val	Ala Pro Asp Ser Ile	Ser Tyr Ile Glu Ala	
2910	2915	2920	
cat ggc acc ggc acc	gta cta ggc gat gca	gtg gag ttg ggt gct	10034
His Gly Thr Gly Thr	Val Leu Gly Asp Ala	Val Glu Leu Gly Ala	
2925	2930	2935	
ttg aat aaa gtg ttc	gac aag cgc gcg gcg	cca tgc ccg atc ggc	10079
Leu Asn Lys Val Phe	Asp Lys Arg Ala Ala	Pro Cys Pro Ile Gly	
2940	2945	2950	
tcg ctg aag gcg aac	atc ggc cat gcc gaa	agc gcc gcg ggc atc	10124
Ser Leu Lys Ala Asn	Ile Gly His Ala Glu	Ser Ala Ala Gly Ile	
2955	2960	2965	
gcc ggc ctg gcc aag	ctg gta ttg cag ttc	agg cat ggc gag ttg	10169
Ala Gly Leu Ala Lys	Leu Val Leu Gln Phe	Arg His Gly Glu Leu	
2970	2975	2980	
gtg cct agt ctg aat	gcg ttt ccc ttg aat	ccc tat att gag ttc	10214
Val Pro Ser Leu Asn	Ala Phe Pro Leu Asn	Pro Tyr Ile Glu Phe	
2985	2990	2995	
ggc cgc ttc cag gta	caa cag cag ccg gca	ccg tgg ccg cgc cgt	10259
Gly Arg Phe Gln Val	Gln Gln Gln Pro Ala	Pro Trp Pro Arg Arg	
3000	3005	3010	
ggc gcc cag ccg ccg	cgc gcc ggg tta tct	gcc ttc ggt gct ggc	10304
Gly Ala Gln Pro Arg	Arg Ala Gly Leu Ser	Ala Phe Gly Ala Gly	
3015	3020	3025	
gga tcg aat gcg cac	cta gtg gta gag gaa	gct ccg gct atg gct	10349

Gly Ser Asn Ala His	Leu Val Val Glu Glu	Ala Pro Ala Met Ala	
3030	3035	3040	
ccc ggg gtc tgc atc	agc gcc agc tct cca	gcc ttg atc gtg ctt	10394
Pro Gly Val Ser Ile	Ser Ala Ser Ser Pro	Ala Leu Ile Val Leu	
3045	3050	3055	
tcg gcg cga acg ctg	cct gcc ttg caa cag	cgt gct cgc gat ctg	10439
Ser Ala Arg Thr Leu	Pro Ala Leu Gln Gln	Arg Ala Arg Asp Leu	
3060	3065	3070	
ctc gtc tgg atg caa	gcg cgg cag gtg gat	gac gtc atg ctg gcc	10484
Leu Val Trp Met Gln	Ala Arg Gln Val Asp	Asp Val Met Leu Ala	
3075	3080	3085	
gac gtt gct tat acg	ctg cac ttg ggc cgc	gtc gcg atg gag caa	10529
Asp Val Ala Tyr Thr	Leu His Leu Gly Arg	Val Ala Met Glu Gln	
3090	3095	3100	
cgc ctg gct ttt acc	gct ggc tcg gct gcc	gag ttg agc gag aaa	10574
Arg Leu Ala Phe Thr	Ala Gly Ser Ala Ala	Glu Leu Ser Glu Lys	
3105	3110	3115	
tta cag gct tac ctg	ggc cat gcg att cgg	gcc gac atc tat ctg	10619
Leu Gln Ala Tyr Leu	Gly His Ala Ile Arg	Ala Asp Ile Tyr Leu	
3120	3125	3130	
agc gag gac acg ccc	ggc aaa ccg gca gcc	gct ccg atc gtg gcc	10664
Ser Glu Asp Thr Pro	Gly Lys Pro Ala Gly	Ala Pro Ile Val Ala	
3135	3140	3145	
gag gaa gat ctg ctc	acg ctg atg gat gcc	tgg atc gaa aag ggc	10709
Glu Glu Asp Leu Leu	Thr Leu Met Asp Ala	Trp Ile Glu Lys Gly	
3150	3155	3160	
cag tac ggt cgt ttg	ctg gag tac tgg acc	aag ggc caa ccg atc	10754
Gln Tyr Gly Arg Leu	Leu Glu Tyr Trp Thr	Lys Gly Gln Pro Ile	
3165	3170	3175	
gac tgg aac aaa ctc	tat tgg cgc aag ctg	tat gcg gac gga cgg	10799
Asp Trp Asn Lys Leu	Tyr Trp Arg Lys Leu	Tyr Ala Asp Gly Arg	
3180	3185	3190	
ccg cgg cgg atc agc	ctg ccc acc tat ccg	ttc gag cac cgg cgt	10844
Pro Arg Arg Ile Ser	Leu Pro Thr Tyr Pro	Phe Glu His Arg Arg	
3195	3200	3205	
tat tgg caa acg ccg	gtg ccg ggc gag cga	agc ctg cac gcc acc	10889
Tyr Trp Gln Thr Pro	Val Pro Gly Glu Arg	Ser Leu His Ala Thr	
3210	3215	3220	
gcg cca gct act cgg	gaa acg gtt gcg gtt	ggt gcc atg ccg gat	10934
Ala Pro Ala Thr Arg	Glu Thr Val Ala Val	Gly Ala Met Pro Asp	
3225	3230	3235	
ccg gcc ggc gct acg	gtg caa gcc cgg ttg	tgc gcc ttg tgc caa	10979
Pro Ala Gly Ala Thr	Val Gln Ala Arg Leu	Cys Ala Leu Cys Gln	
3240	3245	3250	
gtg ttg ttg ggc aaa	ccg gtc acg gcc cag	atg gat ttc ttt gcc	11024
Val Leu Leu Gly Lys	Pro Val Thr Ala Gln	Met Asp Phe Phe Ala	
3255	3260	3265	

gtc ggc ggc cat tcg	gtg ctg gcg atc caa	ttg gtc tcg cgc atc	11069
Val Gly Gly His Ser	Val Leu Ala Ile Gln	Leu Val Ser Arg Ile	
3270	3275	3280	
cgc aaa agc ttc ggg	gtg gag tat ccg gtc	agc gct ttg ttc gaa	11114
Arg Lys Ser Phe Gly	Val Glu Tyr Pro Val	Ser Ala Leu Phe Glu	
3285	3290	3295	
tcg gcg ctg ttg tcg	gac atg gcg cgg cag	atc gaa caa ttg cgg	11159
Ser Ala Leu Leu Ser	Asp Met Ala Arg Gln	Ile Glu Gln Leu Arg	
3300	3305	3310	
gtg aac gga gtc gcc	aag cgc atg ccg gcg	ttg ttg cct gcc ggg	11204
Val Asn Gly Val Ala	Lys Arg Met Pro Ala	Leu Leu Pro Ala Gly	
3315	3320	3325	
cgc gtg ggc gcg att	cct gcg act tat gca	cag gag cgc cta tgg	11249
Arg Val Gly Ala Ile	Pro Ala Thr Tyr Ala	Gln Glu Arg Leu Trp	
3330	3335	3340	
ctc gtc cac gaa cat	atg agt gag caa cgc	agt agt tac aac atc	11294
Leu Val His Glu His	Met Ser Glu Gln Arg	Ser Ser Tyr Asn Ile	
3345	3350	3355	
acc ttt gcc atg cac	ttc aga ggc gtc gac	ttc cgt gct gaa gcg	11339
Thr Phe Ala Met His	Phe Arg Gly Val Asp	Phe Arg Ala Glu Ala	
3360	3365	3370	
atg cgt gcc gca ttg	aac gcg ctg gtg gtg	cgg cac gaa gtg ctg	11384
Met Arg Ala Ala Leu	Asn Ala Leu Val Val	Arg His Glu Val Leu	
3375	3380	3385	
cgc aca cgc ttt ctt	tcg gag gac ggg cag	ctg caa cag gtg atc	11429
Arg Thr Arg Phe Leu	Ser Glu Asp Gly Gln	Leu Gln Gln Val Ile	
3390	3395	3400	
gct gct tcg ttg acg	ttg gag gtg ccg gta	aga gag atg tcg gtc	11474
Ala Ala Ser Leu Thr	Leu Glu Val Pro Val	Arg Glu Met Ser Val	
3405	3410	3415	
gag gag gtc gac ctg	ctg ctg gcc gcg agc	acg cgg gag act ttc	11519
Glu Glu Val Asp Leu	Leu Leu Ala Ala Ser	Thr Arg Glu Thr Phe	
3420	3425	3430	
gat ctg cgg cag ggg	ccc ttg ttc aag gca	cgc atc ctg cgc gtg	11564
Asp Leu Arg Gln Gly	Pro Leu Phe Lys Ala	Arg Ile Leu Arg Val	
3435	3440	3445	
gcg gcc gat cac cat	gtg gtg ttg agc agc	atc cac cac atc att	11609
Ala Ala Asp His His	Val Val Leu Ser Ser	Ile His His Ile Ile	
3450	3455	3460	
tcc gac ggc tgg tcg	ctg gga gtg ttc aac	cgt gac ctg cac cag	11654
Ser Asp Gly Trp Ser	Leu Gly Val Phe Asn	Arg Asp Leu His Gln	
3465	3470	3475	
ctg tac gag gcg tgt	ttg cgc ggc acg ccc	ccc aca ctg ccg acg	11699
Leu Tyr Glu Ala Cys	Leu Arg Gly Thr Pro	Pro Thr Leu Pro Thr	
3480	3485	3490	
ctg gcg gtg cag tat	gcc gac tac gcg ctg	tgg caa cgg caa tgg	11744
Leu Ala Val Gln Tyr	Ala Asp Tyr Ala Leu	Trp Gln Arg Gln Trp	
3495	3500	3505	

gag ctg gcg gct ccg	ctg tcg tac tgg acg	cgg gca ctg gaa ggc	11789
Glu Leu Ala Ala Pro	Leu Ser Tyr Trp Thr	Arg Ala Leu Glu Gly	
3510	3515	3520	
tac gac gac ggc ctg	gac ttg ccc tac gac	cgg ccg cgc ggc gcc	11834
Tyr Asp Asp Gly Leu	Asp Leu Pro Tyr Asp	Arg Pro Arg Gly Ala	
3525	3530	3535	
acg cgg gcg tgg cgg	gca ggg ctg gtc aaa	cac cgc tat ccg ccg	11879
Thr Arg Ala Trp Arg	Ala Gly Leu Val Lys	His Arg Tyr Pro Pro	
3540	3545	3550	
caa ctg gcc cag cag	ttg gcg gcc tac agc	caa cag tac caa gcg	11924
Gln Leu Ala Gln Gln	Leu Ala Ala Tyr Ser	Gln Gln Tyr Gln Ala	
3555	3560	3565	
acg ctg ttc atg agc	ctg ctg gca ggc ctg	gcg ttg gtg ctg ggc	11969
Thr Leu Phe Met Ser	Leu Leu Ala Gly Leu	Ala Leu Val Leu Gly	
3570	3575	3580	
cgt tac gcc gat cgc	aag gac gtg tgc atc	ggc gcg acg gtc tcc	12014
Arg Tyr Ala Asp Arg	Lys Asp Val Cys Ile	Gly Ala Thr Val Ser	
3585	3590	3595	
ggc cgc gac cag ctg	gag ctg gaa gag ctg	atc ggc ttt ttc atc	12059
Gly Arg Asp Gln Leu	Glu Leu Glu Glu Leu	Ile Gly Phe Phe Ile	
3600	3605	3610	
aat att ttg ccg ctg	cgg gtg gac ctg tcg	ggg gat ccg tgc ctg	12104
Asn Ile Leu Pro Leu	Arg Val Asp Leu Ser	Gly Asp Pro Cys Leu	
3615	3620	3625	
gag gag gtg ctg ctg	cgc acg cgt caa gtg	gta ctg gat ggc ttc	12149
Glu Glu Val Leu Leu	Arg Thr Arg Gln Val	Val Leu Asp Gly Phe	
3630	3635	3640	
gcg cac cag tcg gtg	ccg ttc gag cac gtg	ttg cag gcg ctg cgg	12194
Ala His Gln Ser Val	Pro Phe Glu His Val	Leu Gln Ala Leu Arg	
3645	3650	3655	
cgt cag cgc gac agt	agc cag atc ccg ctg	gtg ccg gtg atg ctg	12239
Arg Gln Arg Asp Ser	Ser Gln Ile Pro Leu	Val Pro Val Met Leu	
3660	3665	3670	
cga cac cag aac ttc	ccg acg cag gag att	ggc gat tgg ccc gag	12284
Arg His Gln Asn Phe	Pro Thr Gln Glu Ile	Gly Asp Trp Pro Glu	
3675	3680	3685	
gga gtg cgg ctg acg	cag atg gag ctg ggg	ctg gac cgt agc acg	12329
Gly Val Arg Leu Thr	Gln Met Glu Leu Gly	Leu Asp Arg Ser Thr	
3690	3695	3700	
ccg agc gag ctg gat	tgg cag ttc tac ggc	gac ggc agc tcg ctg	12374
Pro Ser Glu Leu Asp	Trp Gln Phe Tyr Gly	Asp Gly Ser Ser Leu	
3705	3710	3715	
gag ctg acg ctg gaa	tac gcg cag gac ctc	ttc gac gaa gcg acg	12419
Glu Leu Thr Leu Glu	Tyr Ala Gln Asp Leu	Phe Asp Glu Ala Thr	
3720	3725	3730	
gtg cgg cgg atg atc	gca cac cac cag cag	gcg ttg gag gcg atg	12464
Val Arg Arg Met Ile	Ala His His Gln Gln	Ala Leu Glu Ala Met	
3735	3740	3745	

gtg agc cgg cca cag	ctg cgg gtg ggc aag	tgg gac atg ctg acg	12509
Val Ser Arg Pro Gln	Leu Arg Val Gly Lys	Trp Asp Met Leu Thr	
3750	3755	3760	
gcc gaa gag cgc cgg	ctg ttt gcc gcg cta	aat gcg aca ggt acg	12554
Ala Glu Glu Arg Arg	Leu Phe Ala Ala Leu	Asn Ala Thr Gly Thr	
3765	3770	3775	
cca cgg gag tgg ccc	agt ctg gcg cag cag	ttc gaa cgg cag gcg	12599
Pro Arg Glu Trp Pro	Ser Leu Ala Gln Gln	Phe Glu Arg Gln Ala	
3780	3785	3790	
cag gcg acg ccg cag	gcc ata gca tgc gtg	agc gat ggg cag tcg	12644
Gln Ala Thr Pro Gln	Ala Ile Ala Cys Val	Ser Asp Gly Gln Ser	
3795	3800	3805	
tgg agc tat gcg cag	ttg gag gcg cgc gcc	aac cag ctg gca cag	12689
Trp Ser Tyr Ala Gln	Leu Glu Ala Arg Ala	Asn Gln Leu Ala Gln	
3810	3815	3820	
gcg ctg cgt ggg cag	ggc gcg ggc cgg gac	gtg cgg gtg gcg gta	12734
Ala Leu Arg Gly Gln	Gly Ala Gly Arg Asp	Val Arg Val Ala Val	
3825	3830	3835	
cag agt gcg cgc acg	ccg gaa ctg ctg atg	gcc ttg ctg gcg atc	12779
Gln Ser Ala Arg Thr	Pro Glu Leu Leu Met	Ala Leu Leu Ala Ile	
3840	3845	3850	
ttc aag gcc ggt gca	tgc tat gtg ccg atc	gat ccg gcc tac ccg	12824
Phe Lys Ala Gly Ala	Cys Tyr Val Pro Ile	Asp Pro Ala Tyr Pro	
3855	3860	3865	
gcg gcc tac cgc gag	cag atc ctg gcc gag	gtg cag gtg tcg atc	12869
Ala Ala Tyr Arg Glu	Gln Ile Leu Ala Glu	Val Gln Val Ser Ile	
3870	3875	3880	
gtg ctg gag caa gac	gag ctg gcg ctg gac	gag caa ggg cag ttc	12914
Val Leu Glu Gln Asp	Glu Leu Ala Leu Asp	Glu Gln Gly Gln Phe	
3885	3890	3895	
cac aat ccg cgt tgg	cgc gag caa gcc ccg	acg ccg ctg ggg ctg	12959
His Asn Pro Arg Trp	Arg Glu Gln Ala Pro	Thr Pro Leu Gly Leu	
3900	3905	3910	
agg gaa cat ccg ggc	gac ctg gcg tgc gtg	atg gtg acc tcc ggc	13004
Arg Glu His Pro Gly	Asp Leu Ala Cys Val	Met Val Thr Ser Gly	
3915	3920	3925	
tcg acc ggc cgg ccc	aag ggc gtg atg gtg	ccg tat gcg cag ctg	13049
Ser Thr Gly Arg Pro	Lys Gly Val Met Val	Pro Tyr Ala Gln Leu	
3930	3935	3940	
cac aac tgg ctg cat	gca ggc tgg cag cgt	tct gcg ttc gag gcc	13094
His Asn Trp Leu His	Ala Gly Trp Gln Arg	Ser Ala Phe Glu Ala	
3945	3950	3955	
ggg gag cgg gtg ctg	cag aag acc tcg atc	gcc ttt gcg gtg tcg	13139
Gly Glu Arg Val Leu	Gln Lys Thr Ser Ile	Ala Phe Ala Val Ser	
3960	3965	3970	
gta aag gag ttg cta	agc ggg ctg ctg gcg	ggg gtg gaa cag gtg	13184
Val Lys Glu Leu Leu	Ser Gly Leu Leu Ala	Gly Val Glu Gln Val	

3975					3980					3985					
atg	ctg	ccg	gac	gag	cag	gtg	aag	gac	agc	ctg	gcg	ttg	gcg	cgg	13229
Met	Leu	Pro	Asp	Glu	Gln	Val	Lys	Asp	Ser	Leu	Ala	Leu	Ala	Arg	
				3990					3995					4000	
gcg	att	gag	caa	tgg	cag	gtg	acg	cgg	ctg	tac	cta	gtg	cca	tcg	13274
Ala	Ile	Glu	Gln	Trp	Gln	Val	Thr	Arg	Leu	Tyr	Leu	Val	Pro	Ser	
				4005					4010					4015	
cac	ctg	cag	gcg	ctg	ctg	gac	gcg	acg	caa	gga	cga	gac	ggg	cta	13319
His	Leu	Gln	Ala	Leu	Leu	Asp	Ala	Thr	Gln	Gly	Arg	Asp	Gly	Leu	
				4020					4025					4030	
ctg	cac	tcg	ctg	cgt	cac	gtg	gtg	acg	gcg	ggg	gaa	gcg	ttg	ccg	13364
Leu	His	Ser	Leu	Arg	His	Val	Val	Thr	Ala	Gly	Glu	Ala	Leu	Pro	
				4035					4040					4045	
tct	gcg	gtg	cgc	gaa	acg	gtg	cgg	gtg	cgt	ctg	cca	cag	gtg	cag	13409
Ser	Ala	Val	Arg	Glu	Thr	Val	Arg	Val	Arg	Leu	Pro	Gln	Val	Gln	
				4050					4055					4060	
cta	tgg	aac	aac	tat	ggc	tgc	acg	gaa	ctg	aac	gac	gcg	acc	tac	13454
Leu	Trp	Asn	Asn	Tyr	Gly	Cys	Thr	Glu	Leu	Asn	Asp	Ala	Thr	Tyr	
				4065					4070					4075	
cat	cgg	tcg	gat	acg	gtg	gcg	cca	gga	acg	ttt	gtg	ccg	atc	ggc	13499
His	Arg	Ser	Asp	Thr	Val	Ala	Pro	Gly	Thr	Phe	Val	Pro	Ile	Gly	
				4080					4085					4090	
gca	ccg	atc	gcc	aac	acc	gag	gta	tac	gtg	ctg	gac	cgg	cag	ctg	13544
Ala	Pro	Ile	Ala	Asn	Thr	Glu	Val	Tyr	Val	Leu	Asp	Arg	Gln	Leu	
				4095					4100					4105	
cgg	cag	gtg	ccg	atc	ggg	gtg	atg	ggc	gag	ctg	cac	gta	cac	agc	13589
Arg	Gln	Val	Pro	Ile	Gly	Val	Met	Gly	Glu	Leu	His	Val	His	Ser	
				4110					4115					4120	
gtg	ggg	atg	gcg	cgc	ggc	tac	tgg	aac	cgg	ccg	ggg	ctg	acg	gcc	13634
Val	Gly	Met	Ala	Arg	Gly	Tyr	Trp	Asn	Arg	Pro	Gly	Leu	Thr	Ala	
				4125					4130					4135	
tcg	cgc	ttc	atc	gcg	cac	ccg	tat	agc	gag	gag	ccg	ggc	aca	cgg	13679
Ser	Arg	Phe	Ile	Ala	His	Pro	Tyr	Ser	Glu	Glu	Pro	Gly	Thr	Arg	
				4140					4145					4150	
ctg	tac	aag	acc	ggt	gac	atg	gta	cgc	cgg	ctg	gcg	gac	ggg	acg	13724
Leu	Tyr	Lys	Thr	Gly	Asp	Met	Val	Arg	Arg	Leu	Ala	Asp	Gly	Thr	
				4155					4160					4165	
ctg	gaa	tac	ctg	ggc	cga	cag	gac	ttc	gag	gtc	aag	gtg	cgc	ggc	13769
Leu	Glu	Tyr	Leu	Gly	Arg	Gln	Asp	Phe	Glu	Val	Lys	Val	Arg	Gly	
				4170					4175					4180	
cac	cgg	gtg	gat	acg	cgg	cag	gtg	gag	gcg	gcc	ttg	cgg	gcg	cag	13814
His	Arg	Val	Asp	Thr	Arg	Gln	Val	Glu	Ala	Ala	Leu	Arg	Ala	Gln	
				4185					4190					4195	
ccc	gcg	gtg	gcc	gag	gcg	gtg	gtg	agc	ggt	cac	cgg	gtg	gac	ggg	13859
Pro	Ala	Val	Ala	Glu	Ala	Val	Val	Ser	Gly	His	Arg	Val	Asp	Gly	
				4200					4205					4210	
gac	atg	cag	ttg	gtg	gcc	tat	gtg	gtg	gcg	cgt	gaa	ggg	cag	gca	13904
Asp	Met	Gln	Leu	Val	Ala	Tyr	Val	Val	Ala	Arg	Glu	Gly	Gln	Ala	

4215	4220	4225	
ccg agc gcg ggc gag	ttg aaa caa cag ctg	tcg gcg cag ttg ccg	13949
Pro Ser Ala Gly Glu	Leu Lys Gln Gln Leu	Ser Ala Gln Leu Pro	
4230	4235	4240	
acc tac atg ctg ccg	acc gtg tac cag tgg	ctg gag cag ttg ccg	13994
Thr Tyr Met Leu Pro	Thr Val Tyr Gln Trp	Leu Glu Gln Leu Pro	
4245	4250	4255	
cgg ctg tcc aac ggc	aag ttg gac cgg ttg	gcg ctg ccg gcg ccg	14039
Arg Leu Ser Asn Gly	Lys Leu Asp Arg Leu	Ala Leu Pro Ala Pro	
4260	4265	4270	
cag gtg gta cac gcg	cag gag tac gtc gcg	cca cgc aac gag gcc	14084
Gln Val Val His Ala	Gln Glu Tyr Val Ala	Pro Arg Asn Glu Ala	
4275	4280	4285	
gag caa cgg ctg gcg	gca ctg ttt gcc gag	gtg ctg cgg gtg gag	14129
Glu Gln Arg Leu Ala	Ala Leu Phe Ala Glu	Val Leu Arg Val Glu	
4290	4295	4300	
cag gtg ggc atc cac	gac aac ttc ttc gcc	ttg ggt ggg cac tcg	14174
Gln Val Gly Ile His	Asp Asn Phe Phe Ala	Leu Gly Gly His Ser	
4305	4310	4315	
ctg tct gca tcg caa	ctg atc tcg cgc atc	cgc caa agt ttt cac	14219
Leu Ser Ala Ser Gln	Leu Ile Ser Arg Ile	Arg Gln Ser Phe His	
4320	4325	4330	
gtc gat ctg ccg ctg	agc cgg atc ttc gag	gca ccc acg atc gag	14264
Val Asp Leu Pro Leu	Ser Arg Ile Phe Glu	Ala Pro Thr Ile Glu	
4335	4340	4345	
ggc ctg gtc agg cag	cta gcg ttg cct agt	gaa ggc ggc gtg gcc	14309
Gly Leu Val Arg Gln	Leu Ala Leu Pro Ser	Glu Gly Gly Val Ala	
4350	4355	4360	
agc atc gcc agg gta	gcg cga aac cgg acg	atc cca ttg tcg ctg	14354
Ser Ile Ala Arg Val	Ala Arg Asn Arg Thr	Ile Pro Leu Ser Leu	
4365	4370	4375	
ttc cag gaa cgc ctg	tgg ttc gtg cac caa	cac atg cct gag caa	14399
Phe Gln Glu Arg Leu	Trp Phe Val His Gln	His Met Pro Glu Gln	
4380	4385	4390	
cgc acc agt tac aac	ggc acg ctc gcc ttg	cgt ttg cgt ggt cct	14444
Arg Thr Ser Tyr Asn	Gly Thr Leu Ala Leu	Arg Leu Arg Gly Pro	
4395	4400	4405	
ttg tcg gtg gaa gcg	atg cgt gca gcg ctg	cgt gcg tta gtg ctg	14489
Leu Ser Val Glu Ala	Met Arg Ala Ala Leu	Arg Ala Leu Val Leu	
4410	4415	4420	
cgc cac gaa atc ttg	cgt acc cgc ttc gtg	ttg ccg acc ggt gct	14534
Arg His Glu Ile Leu	Arg Thr Arg Phe Val	Leu Pro Thr Gly Ala	
4425	4430	4435	
agc gag ccg gtg cag	gtc att gac gag cac	agc gat ttc cag ctc	14579
Ser Glu Pro Val Gln	Val Ile Asp Glu His	Ser Asp Phe Gln Leu	
4440	4445	4450	
tca gta cag cta gtc	gag gat act gag atc	gcg tcg ctg atg gat	14624

Ser Val Gln Leu Val	Glu Asp Thr Glu Ile	Ala Ser Leu Met Asp	
4455	4460	4465	
gaa ctg gca agt cat	atc tac gac tta gcc	aac ggc ccg ctg ttc	14669
Glu Leu Ala Ser His	Ile Tyr Asp Leu Ala	Asn Gly Pro Leu Phe	
4470	4475	4480	
att gca tgc ctt ttg	caa ctg gat gag caa	gaa cat gtg ctg cta	14714
Ile Ala Cys Leu Leu	Gln Leu Asp Glu Gln	Glu His Val Leu Leu	
4485	4490	4495	
atc ggc atg cat cac	ctt atc tac gac gct	tgg tcg caa ttc acc	14759
Ile Gly Met His His	Leu Ile Tyr Asp Ala	Trp Ser Gln Phe Thr	
4500	4505	4510	
gtg atg aac cgc gat	cta cgc gtg ctg tat	cac cgc cac ctc gga	14804
Val Met Asn Arg Asp	Leu Arg Val Leu Tyr	His Arg His Leu Gly	
4515	4520	4525	
ctt gcc ggc gga gat	ctg ccg gaa tta ccg	atc caa tat gcc gac	14849
Leu Ala Gly Gly Asp	Leu Pro Glu Leu Pro	Ile Gln Tyr Ala Asp	
4530	4535	4540	
tat gcg atc tgg caa	cgc gcc cag aac ctg	gac gcg caa ctg gcc	14894
Tyr Ala Ile Trp Gln	Arg Ala Gln Asn Leu	Asp Ala Gln Leu Ala	
4545	4550	4555	
tat tgg cag gct atg	ttg cac gac tac gac	gac ggc ctg gag ctg	14939
Tyr Trp Gln Ala Met	Leu His Asp Tyr Asp	Asp Gly Leu Glu Leu	
4560	4565	4570	
ccc tac gac tat ccg	cgt ccg cgc aat cgc	acc tgg cac gca gcg	14984
Pro Tyr Asp Tyr Pro	Arg Pro Arg Asn Arg	Thr Trp His Ala Ala	
4575	4580	4585	
gtc tac aca cac acc	tat ccg gct gaa ctg	gta cag cgc ttt gcc	15029
Val Tyr Thr His Thr	Tyr Pro Ala Glu Leu	Val Gln Arg Phe Ala	
4590	4595	4600	
ggc ttc gta cag gcg	cat cag tcg acc ttg	ttc atc ggg ctg ttg	15074
Gly Phe Val Gln Ala	His Gln Ser Thr Leu	Phe Ile Gly Leu Leu	
4605	4610	4615	
gcc agc ttc gcg gtc	gtg ttg aac aaa tac	acc ggc cgg gac gac	15119
Ala Ser Phe Ala Val	Val Leu Asn Lys Tyr	Thr Gly Arg Asp Asp	
4620	4625	4630	
ttg tgc atc ggt acc	acc acg gca ggg cgc	acg cac ctg gag ctg	15164
Leu Cys Ile Gly Thr	Thr Thr Ala Gly Arg	Thr His Leu Glu Leu	
4635	4640	4645	
gag aac ctg atc ggt	ttc ttc atc aac atc	ttg cct ttg cgc ttg	15209
Glu Asn Leu Ile Gly	Phe Phe Ile Asn Ile	Leu Pro Leu Arg Leu	
4650	4655	4660	
cgc ttg gac ggc gat	ccg gac gtt gcc gaa	atc atg cgg cga aca	15254
Arg Leu Asp Gly Asp	Pro Asp Val Ala Glu	Ile Met Arg Arg Thr	
4665	4670	4675	
cgg ttg gtg gcg atg	agc gcg ttt gag aac	cag gcg cta ccg ttc	15299
Arg Leu Val Ala Met	Ser Ala Phe Glu Asn	Gln Ala Leu Pro Phe	
4680	4685	4690	

gag cac ctg ctc aac gcc ctg cac aag caa cgt gac acc agc cgg 15344
 Glu His Leu Leu Asn Ala Leu His Lys Gln Arg Asp Thr Ser Arg
 4695 4700 4705

 att ccg cta gtt ccg gtg gtg atg cgt cat cag aac ttc ccg gac 15389
 Ile Pro Leu Val Pro Val Val Met Arg His Gln Asn Phe Pro Asp
 4710 4715 4720

 acg atc ggc gac tgg agc gat ggc atc cgt acc gaa gtg atc cag 15434
 Thr Ile Gly Asp Trp Ser Asp Gly Ile Arg Thr Glu Val Ile Gln
 4725 4730 4735

 cgc gat ctg cgt gcc acc ccc aat gaa atg gac ctg caa ttc ttc 15479
 Arg Asp Leu Arg Ala Thr Pro Asn Glu Met Asp Leu Gln Phe Phe
 4740 4745 4750

 ggc gac ggt acg ggg ctt tcg gtc aca gtg gaa tac gcg gcg gag 15524
 Gly Asp Gly Thr Gly Leu Ser Val Thr Val Glu Tyr Ala Ala Glu
 4755 4760 4765

 ctg ttc tca gaa gcg acc att cgc cgc ctg atc cac cat cac caa 15569
 Leu Phe Ser Glu Ala Thr Ile Arg Arg Leu Ile His His His Gln
 4770 4775 4780
 ctc gtc ctg gag cag atg ttg gcg gcc cat gaa agc gcc acg tgc 15614
 Leu Val Leu Glu Gln Met Leu Ala Ala His Glu Ser Ala Thr Cys
 4785 4790 4795

 ccc ttg gat gtt gcc gac tagcaaaagc cgcccgccgt caccggttca 15662
 Pro Leu Asp Val Ala Asp
 4800

 tcgatagcga gggcaatcat ggattcagcg ttacctacat ctgcatttac cttcgatctc 15722
 ttttacacca cggttaacgc ctactatcgc actgccgcag tcaaggcggc gatcgaactg 15782
 gggctattcg atgtggtggg gcgscagggc cgaactcccg cagccatcgc cgaggcctgc 15842
 caggcgctgc cgcgcggcat tcgcatcctt tgctattacc tagtatcgat cggttttcta 15902
 cgccgcaacg gtggcctgtt ctacatagat cgcaacatgg ccatgtacct ggatcgtagt 15962
 tcgcccggct acctgggtgg cagcatcaag ttcctgctct cgccctacat catgagcgcc 16022
 ttcaccgatc tgaccgccgt agtcaggacc ggcaagatca acctggcgca ggacggcgtg 16082
 gtggcaccgg atcaccgcga gtgggtggaa tttgcacgcg cgatggcacc gatgatggcg 16142
 ctgccctcgg cgttgatcgc caatatggtg tcgttgcccc ctgatcggcc gattcgtgtg 16202
 ctggacgtgg cagccggcca cggcctgttc ggcatcgctt tcgcgcagcg cttccgccag 16262
 gctgaagtga gcttcctgga ctgggacaac gtgctagacg tagcacgcga aaacgcccag 16322
 gcggccaaag tggccgagcg agcgcgtttc ctgcccggca acgcattcga cctcgattar 16382
 ggcagcggt acgacgtgat cttgttgacc aacttcctgc accatttcga tgaggctgat 16442
 ggcgagcgca tcttggttaa gacgcgcgat gcgctgaacg acgacggcat ggtgatcact 16502
 ttcgaattc 16511

<211> 4801
 <212> PRT
 <213> Xanthomonas albilineans

<400> 2

Leu Pro Asn Ala Leu Met Gln Ile Thr Leu Val Ala Val Gln Phe Ala
 1 5 10 15

Gly Val Leu Leu Gly Val Thr Ala Arg Ala Ala Ile Pro Asn Lys Ala
 20 25 30

Gly Met Arg Arg Ala Trp Pro Pro Phe Pro Gln Ala Cys Cys Arg Ser
 35 40 45

Ile Ala Tyr Leu Met Gln Arg Ser Pro Met Ser Pro Leu Gln Gln Thr
 50 55 60

Leu Leu Thr Arg Leu Ala Ser Ala Ala Ala Ser Arg Thr Met Ile Glu
 65 70 75 80

Phe Pro Arg Pro Glu His Ala Ser Pro Gln Cys Cys Asp Asp Ala Glu
 85 90 95

Leu Ala Arg Leu Ile Val Gln Leu Ser Ala Gly Leu Gln Pro Leu Ala
 100 105 110

Met Pro Gly Thr Tyr Val Ile Ile Ala Ala Pro His Gly Gly Leu Phe
 115 120 125

Ala Ala Ala Leu Leu Ala Cys Leu His Ala Asn Leu Val Ala Val Pro
 130 135 140

Phe Pro Leu Asp Val Ala Gln Pro Asn Glu Arg Glu Gln Ala Arg Leu
 145 150 155 160

Glu Thr Ile His Ala Gln Leu Met Glu His Gly Asn Val Ala Val Leu
 165 170 175

Leu Asp Asp Val Ala Asp Arg Ser Ala Phe Ala Arg Met Ala His Ala
 180 185 190

Ala Gly Thr Phe Leu Ala Thr Phe Ala Asp Leu Lys Arg Glu Ser Thr
 195 200 205

Ser Ala Ser Leu Cys Pro Ala Ser Pro Ser Asp Ala Ala Leu Leu Leu
 210 215 220

Phe Thr Ser Gly Ser Ser Gly Glu Ser Lys Gly Ile Leu Leu Ser His
 225 230 235 240

Arg Asn Leu His His Gln Ile Gln Ala Gly Ile Arg Gln Trp Ser Leu
245 250 255

Asp Glu His Ser His Val Val Thr Trp Leu Ser Pro Ala His Asn Phe
260 265 270

Gly Leu His Phe Gly Leu Leu Ala Pro Trp Phe Ser Gly Ala Thr Val
275 280 285

Ser Phe Ile His Pro His Ser Tyr Met Lys Arg Pro Gly Phe Trp Leu
290 295 300

Glu Thr Val Ala Ala Arg Asp Ala Thr His Met Ala Ala Pro Asn Phe
305 310 315 320

Ala Phe Asp Tyr Cys Cys Asp Trp Val Met Val Glu Gln Leu Pro Pro
325 330 335

Ser Ala Leu Ser Thr Leu Thr His Ile Val Cys Gly Gly Glu Pro Val
340 345 350

Arg Ala Ser Thr Met Gln Arg Phe Phe Glu Lys Phe Ala Gly Leu Gly
355 360 365

Ala Arg Thr Gln Thr Phe Met Pro His Phe Gly Leu Ser Glu Thr Gly
370 375 380

Ala Leu Ser Thr Leu Asp Glu Ala Pro Gln Gln Arg Val Leu Glu Leu
385 390 395 400

Asp Ala Asp Ala Leu Asn Lys Arg Lys Arg Val Ala Ala Gly Ala Ser
405 410 415

Gln Ala Arg Val Thr Val Leu Asn Cys Gly Ala Val Asp Gln Asp Val
420 425 430

Glu Leu Arg Ile Val Cys Pro Glu Gly Glu Thr Leu Cys Arg Pro Asp
435 440 445

Glu Ile Gly Glu Ile Trp Val Lys Ser Pro Ala Ile Ala Arg Gly Tyr
450 455 460

Leu Phe Ala Lys Pro Ala Asp Gln Arg Gln Phe Asn Cys Ser Ile Arg
465 470 475 480

His Thr Asp Asp Ser Gly Tyr Phe Arg Thr Gly Asp Leu Gly Phe Ile

485										490										495																																		
Ala	Asp	Gly	Cys	Leu	Tyr	Val	Thr	Gly	Arg	Val	Lys	Glu	Val	Leu	Ile																																							
				500						505					510																																							
Ile	Arg	Gly	Lys	Asn	His	Tyr	Pro	Ala	His	Ile	Glu	Ala	Ser	Ile	Ala																																							
				515				520						525																																								
Ala	Thr	Ala	Ser	Pro	Gly	Ala	Leu	Met	Pro	Val	Val	Phe	Ser	Ile	Glu																																							
				530				535						540																																								
Arg	Gln	Asp	Glu	Glu	Arg	Val	Ala	Ala	Val	Ile	Ala	Val	Asn	His	Pro																																							
						550					555				560																																							
Trp	Thr	Pro	Ala	Ala	Cys	Ala	Ala	Gln	Ala	His	Lys	Ile	Arg	Gln	Gln																																							
					565					570					575																																							
Val	Ala	Asp	Gln	His	Gly	Val	Ala	Leu	Ala	Glu	Leu	Ala	Phe	Ala	Glu																																							
				580					585					590																																								
His	Arg	His	Val	Phe	Gly	Thr	Tyr	Pro	Gly	Lys	Leu	Lys	Arg	Arg	Leu																																							
				595				600						605																																								
Val	Lys	Glu	Ala	Tyr	Val	Asn	Gly	Gln	Leu	Pro	Leu	Leu	Trp	His	Glu																																							
				610				615					620																																									
Gly	Lys	Asn	Arg	Asp	Val	Pro	Ala	Ala	Ala	Ala	Asp	Asp	Arg	Gln	Ala																																							
				625				630					635		640																																							
Gln	His	Val	Ala	Asp	Leu	Cys	Arg	Lys	Val	Phe	Leu	Pro	Val	Leu	Gly																																							
					645					650				655																																								
Val	Ala	Pro	Pro	His	Ala	Gln	Trp	Pro	Leu	Cys	Glu	Leu	Ala	Leu	Asp																																							
					660					665				670																																								
Ser	Leu	Gln	Cys	Val	Arg	Leu	Ala	Gly	Ala	Ile	Glu	Glu	Cys	Tyr	Gly																																							
					675					680				685																																								
Val	Pro	Phe	Glu	Pro	Thr	Leu	Leu	Phe	Lys	Leu	Glu	Thr	Val	Gly	Ala																																							
					690					695				700																																								
Ile	Ala	Glu	Tyr	Val	Leu	Ala	His	Gly	Arg	Gln	Ala	Pro	Thr	Pro	Thr																																							
					710					715				720																																								
Arg	Ala	Pro	Val	Ala	Ser	Thr	Thr	Cys	Ser	Glu	Glu	Pro	Ile	Ala	Ile																																							
					725					730				735																																								
Val	Ala	Met	His	Cys	Glu	Val	Pro	Gly	Ala	Gly	Glu	Asn	Thr	Glu	Ala																																							

740	745	750
Leu Trp Ser Phe Leu Arg Ser Asp Val Asn Ala Ile Arg Pro Ile Glu		
755	760	765
Ser Thr Arg Pro Asp Leu Trp Ala Ala Met Arg Ala Tyr Pro Gly Leu		
770	775	780
Ala Gly Glu Gln Leu Pro Arg Tyr Ala Gly Phe Leu Asp Asp Val Asp		
785	790	795
Ala Phe Asp Ala Ala Phe Phe Gly Ile Ser Arg Arg Glu Ala Glu Cys		
805	810	815
Met Asp Pro Gln Gln Arg Lys Val Leu Glu Met Val Trp Lys Leu Ile		
820	825	830
Glu Gln Ala Gly His Asp Pro Leu Ser Trp Gly Gly Gln Pro Val Gly		
835	840	845
Leu Phe Val Gly Ala His Thr Ser Asp Tyr Gly Glu Leu Leu Ala Ser		
850	855	860
Gln Pro Gln Leu Met Ala Gln Cys Gly Ala Tyr Ile Asp Ser Gly Ser		
865	870	875
His Leu Thr Met Ile Pro Asn Arg Ala Ser Arg Trp Phe Asn Phe Thr		
885	890	895
Gly Pro Ser Glu Val Ile Asn Ser Ala Cys Ser Ser Ser Leu Val Ala		
900	905	910
Leu His Arg Ala Val Gln Ser Leu Arg Gln Gly Glu Ser Ser Val Ala		
915	920	925
Leu Val Leu Gly Val Asn Leu Ile Leu Ala Pro Lys Val Leu Leu Ala		
930	935	940
Ser Ala Ser Ala Gly Met Leu Ser Pro Asp Gly Arg Cys Lys Thr Leu		
945	950	955
Asp Ala Ala Ala Asp Gly Phe Val Arg Ser Glu Gly Ile Ala Gly Val		
965	970	975
Ile Leu Lys Pro Leu Ala Gln Ala Leu Ala Asp Gly Asp Arg Val Tyr		
980	985	990

Gly Leu Val Arg Gly Val Ala Val Asn His Gly Gly Arg Ser Asn Ser
 995 1000 1005

Leu Arg Ala Pro Asn Val Asn Ala Gln Arg Gln Leu Leu Ile Arg
 1010 1015 1020

Thr Tyr Gln Glu Ala Gly Val Glu Pro Ala Ser Val Gly Tyr Val
 1025 1030 1035

Glu Leu His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Ile
 1040 1045 1050

Gln Ala Leu Lys Glu Ala Phe Ile Ala Leu Gly Ala Gln Ala Ala
 1055 1060 1065

Pro Ser Asn Cys Gly Ile Gly Ser Val Lys Ser Ala Leu Gly His
 1070 1075 1080

Leu Glu Ala Ala Ala Gly Leu Thr Gly Leu Ile Lys Val Leu Leu
 1085 1090 1095

Met Leu Lys His Gly Glu Gln Ala Gly Thr Arg His Phe Ser Thr
 1100 1105 1110

Leu Asn Pro Leu Ile Asp Leu Arg Gly Thr Ser Phe Glu Val Val
 1115 1120 1125

Ala Gln His Arg Ala Trp Pro Ser Gln Val Gly Ile His Gly Thr
 1130 1135 1140

Leu Leu Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly Gly
 1145 1150 1155

Ala Asn Ala His Ala Ile Val Glu Glu His Val Ile Ala Thr Pro
 1160 1165 1170

Pro Ser Thr Ser Ser Ala Gly Gly Pro Val Gly Ile Val Leu Ser
 1175 1180 1185

Ala Gly Ser Glu Ala Val Leu Arg Gln Gln Val Leu Ala Leu Ser
 1190 1195 1200

Ala Trp Leu Arg Gln Gln Ser Pro Thr Pro Ala Gln Met Ile Asp
 1205 1210 1215

Val Ala Tyr Thr Leu Gln Val Gly Arg Ala Ala Leu Ser His Arg
 1220 1225 1230

Leu Ala Phe Ser Ala Thr Asp	Ala Glu Gln Ala Leu	Ala Arg Leu
1235	1240	1245
Glu Gly Arg Leu Ala Gly Val	Met Asp Ala Glu Val	His His Gly
1250	1255	1260
Val Val Asp Ala Ala Ala Thr	Ala Pro Glu His Gly	Arg Gln Thr
1265	1270	1275
Arg Glu Gly Leu Ala Gly Leu	Leu Arg Ala Trp Thr	Gln Gly Val
1280	1285	1290
Arg Val Asp Trp Ser Ala Leu	Tyr Gly Ile Gln Arg	Pro Gln Arg
1295	1300	1305
Val Ser Leu Pro Val Tyr Pro	Phe Ala Arg Glu Arg	Tyr Trp Leu
1310	1315	1320
Pro Gly Gln Ala Met His Ala	Ala Ala Asp Ala His	Pro Met Leu
1325	1330	1335
Gln Leu Leu His Ala Asn Ala	Lys Leu His Arg Tyr	Ala Leu Arg
1340	1345	1350
Arg Ser Gly Cys Ala Ser Phe	Leu Val Asp His Cys	Val Asp Gly
1355	1360	1365
Arg Gln Val Leu Pro Ala Ala	Val Gln Leu Glu Leu	Val Arg Ala
1370	1375	1380
Val Ala Gln Arg Val Met Ala	Gln Asp Glu Gly Cys	Ile Glu Leu
1385	1390	1395
Ala Gln Val Ala Phe Leu His	Pro Leu Met Met Glu	Glu Thr Glu
1400	1405	1410
Leu Glu Val Glu Ile Glu Leu	Ser Lys Ser Asp Gln	Asp Glu Phe
1415	1420	1425
Asp Phe Gln Leu His Asp Ala	His Arg Gln Gln Val	Phe Ser Gln
1430	1435	1440
Gly His Val Arg Arg Arg Val	Tyr Thr Ala Thr Pro	Arg Leu Asp
1445	1450	1455
Leu Ala Gln Leu Gln Lys Leu	Cys Ala Glu Arg Val	Leu Ser Gly
1460	1465	1470

Glu Asp	Cys Tyr	Ala His	Phe	Thr Ala	Cys Gly	Leu	Gln Leu	Gly			
1475			1480			1485					
Asp Arg	Leu Lys	Ser Val	Gln	Ser Ile	Gly Cys	Gly	Arg Asn	Gly			
1490			1495			1500					
Glu Gly	Glu Pro	Ile Ala	Leu	Gly Val	Leu Arg	Leu	Pro Pro	Ser			
1505			1510			1515					
Ser Val	Glu Asp	Ser His	Val	Leu Pro	Pro Ser	Leu	Leu Asp	Gly			
1520			1525			1530					
Ala Leu	Gln Cys	Ser Leu	Gly	Leu Gln	Arg Asp	Val	Glu His	Ile			
1535			1540			1545					
Ala Met	Pro Tyr	Thr Leu	Glu	Arg Met	Thr Val	His	Ala Pro	Ile			
1550			1555			1560					
Pro Pro	Glu Ala	Trp Val	Leu	Leu Arg	His Gly	His	Ala Ala	Arg			
1565			1570			1575					
Gln Ser	Leu Asp	Ile Asp	Leu	Leu Asp	Ser Glu	Gly	Arg Val	Cys			
1580			1585			1590					
Val Ser	Leu Gly	Asn Tyr	Thr	Gly Arg	Ala Pro	Lys	Ala Val	Ser			
1595			1600			1605					
Ala Val	Arg Ala	Leu Val	Leu	Ala Pro	Val Trp	Gln	Ala Leu	Thr			
1610			1615			1620					
Glu Thr	Ala Pro	Ala Trp	Pro	Asp Pro	Ala Glu	Arg	Ile Val	Thr			
1625			1630			1635					
Val Gly	Asp Asp	Ala Trp	Arg	Ser His	Phe Gly	Phe	Asp Glu	Pro			
1640			1645			1650					
Ala Leu	Ser Leu	Glu Asp	Ser	Val Glu	Val Ile	Ala	Thr Arg	Leu			
1655			1660			1665					
Gly Gln	Ser Gly	Lys Phe	Asp	His Leu	Val Trp	Ile	Val Pro	Ile			
1670			1675			1680					
Ala Glu	Ser Glu	Thr Asp	Ile	Ala Ala	Gln Gly	Ser	Ala Ala	Ile			
1685			1690			1695					
Ala Gly	Phe Arg	Leu Val	Lys	Ala Leu	Leu Ala	Leu	Gly Tyr	Ala			
1700			1705			1710					

His Arg Pro Leu Gly Leu Thr Val Leu Thr Arg Gln Ala Leu Thr
1715 1720 1725

Arg Gln Pro Ser His Ala Ala Val His Gly Leu Ile Gly Thr Leu
1730 1735 1740

Ala Lys Glu Tyr Cys Asn Trp Lys Ile Arg Leu Leu Asp Leu Pro
1745 1750 1755

Ser Val Lys Ser Trp Pro Gln Trp Glu Gln Leu Arg Ser Leu Pro
1760 1765 1770

Trp His Ala Gln Gly Glu Ala Leu Ile Gly Arg Gly Thr Cys Trp
1775 1780 1785

Tyr Arg Arg Gln Leu Cys Glu Val Leu Pro Leu Pro Ser Leu Glu
1790 1795 1800

Pro Pro Pro Tyr Arg Val Gly Gly Val Tyr Val Val Ile Gly Gly
1805 1810 1815

Ala Gly Gly Leu Gly Glu Val Leu Ser Glu His Leu Ile Arg Thr
1820 1825 1830

Tyr Asp Ala Gln Leu Ile Trp Ile Gly Arg Arg Val Leu Asp Glu
1835 1840 1845

Gly Ile Ala Arg Lys Gln Thr Arg Leu Ala Ser Leu Gly Arg Ala
1850 1855 1860

Pro His Tyr Ile Ser Ala Asp Ala Ser Asp Pro Ala Ala Leu Gln
1865 1870 1875

Ala Ala His Asn Glu Ile Val Ala Leu His Gly Gln Pro His Gly
1880 1885 1890

Leu Ile Leu Ser Asn Ile Val Leu Lys Asp Ala Ser Leu Ala Arg
1895 1900 1905

Met Glu Glu Ala Asp Phe Arg Asp Val Leu Ala Ala Lys Leu Asp
1910 1915 1920

Val Ser Val Cys Ala Ala Gln Val Phe Gly Thr Ala Pro Leu Asp
1925 1930 1935

Phe Val Leu Phe Phe Ser Ser Ile Gln Ser Thr Thr Lys Ala Ala

1940		1945		1950
Gly Gln Gly Asn Tyr Ala Ala Gly Cys Cys Tyr Val Asp Ala Phe				
1955		1960		1965
Gly Glu Leu Trp Ala Arg Arg Gly Leu Arg Val Lys Thr Ile Asn				
1970		1975		1980
Trp Gly Tyr Trp Gly Ser Val Gly Val Val Ala Gly Glu Asp Tyr				
1985		1990		1995
Arg Arg Arg Met Ala Gln Lys His Met Ala Ser Ile Glu Gly Ala				
2000		2005		2010
Glu Ala Met Gln Val Leu Ser Gln Leu Leu Cys Ala Pro Leu Gln				
2015		2020		2025
Arg Leu Ala Tyr Val Lys Ile Asp Asp Ala Asn Ala Met Arg Ala				
2030		2035		2040
Leu Gly Val Val Glu Asp Glu Ser Val Gln Ile Pro Val His Ala				
2045		2050		2055
Pro Ala Glu Pro Pro Arg Gly Gln Pro Gly Pro Val Val Glu Leu				
2060		2065		2070
Ser Val Asn Leu Asp Ala Arg Arg Glu Arg Glu Thr Leu Leu Ala				
2075		2080		2085
Ala Trp Leu Leu Glu Leu Ile Glu Gln Leu Gly Gly Phe Pro Pro				
2090		2095		2100
Ala Ser Phe Asp Ile Ala Thr Leu Ala Gln Arg Leu His Ile Val				
2105		2110		2115
Pro Ala Tyr Arg Ser Trp Leu Glu His Ser Val Arg Met Leu Gly				
2120		2125		2130
Val Tyr Gly Tyr Leu Arg Ala Thr Gly Glu Ser Arg Phe Glu Leu				
2135		2140		2145
Ala Asp Lys Pro Pro Asp Asp Ala Arg Gly Ala Trp Asn Ala His				
2150		2155		2160
Val His Glu Ala Ser Val Glu Ala Gly Glu Glu Ala Gln Arg Arg				
2165		2170		2175
Leu Leu Asp Arg Cys Met Arg Ala Leu Pro Ala Val Leu Arg Gly				

2180	2185	2190
Glu Arg Lys Ala Thr Glu Leu Leu Phe Pro Glu Gly Ser Met Ala 2195 2200 2205		
Trp Val Glu Gly Ile Tyr Gln Asn Asn Pro Leu Ala Asp Tyr Phe 2210 2215 2220		
Asn Ala Gln Leu Val Thr Arg Leu Ile Ala Tyr Leu Arg Arg Arg 2225 2230 2235		
Leu Glu Ser Thr Pro Thr Ala Arg Leu Lys Leu Cys Glu Ile Gly 2240 2245 2250		
Ala Gly Ser Gly Gly Thr Thr Ala Ser Val Leu Gln Gln Leu Gln 2255 2260 2265		
Ala Tyr Gly Glu His Ile Glu Glu Tyr Leu Tyr Thr Asp Leu Ser 2270 2275 2280		
Pro Val Phe Leu His His Ala Glu Lys His Tyr Gln Pro Arg Ala 2285 2290 2295		
Pro Tyr Leu Arg Thr Ala Cys Phe Asp Val Ala Arg Ala Pro Thr 2300 2305 2310		
Ala Gln Ala Leu Glu Ser Gly Gly Tyr Asp Val Val Ile Ala Ala 2315 2320 2325		
Asn Val Leu His Ala Thr Arg Asp Ile Ala Lys Thr Leu Arg Asn 2330 2335 2340		
Ala Lys Ala Leu Leu Lys Pro Gly Gly Leu Leu Leu Leu Asn Glu 2345 2350 2355		
Val Ile Glu Arg Ser Leu Val Leu His Leu Thr Phe Gly Leu Leu 2360 2365 2370		
Glu Ser Trp Trp Leu Pro Gln Asp Lys Ile Leu Arg Leu Ala Gly 2375 2380 2385		
Ser Pro Leu Leu Ala Cys Ala Thr Trp Arg Ser Leu Leu Glu Ala 2390 2395 2400		
Glu Gly Phe Ala Gly Leu Ser Val His Arg Ala Gln Pro Asp Ala 2405 2410 2415		

Gly Gln	Ala Ile Ile Cys	Ala Tyr Ser Asp	Gly Ile	Val Arg Gln
2420		2425	2430	
Ala Ser	Thr Ile Glu Val	Ala Arg Asn Glu Lys	Val Thr Val Pro	
2435		2440	2445	
Ser Gln	Pro Ala Glu Ala	Gly Glu Ser Pro Leu	Asp Leu Val Lys	
2450		2455	2460	
Lys Leu	Leu Gly Arg Ile	Leu Lys Met Asp Pro	Ala Thr Leu Asp	
2465		2470	2475	
Thr Ser	His Pro Leu Glu	Tyr Tyr Gly Val Asp	Ser Ile Val Ala	
2480		2485	2490	
Ile Glu	Leu Ala Met Ala	Leu Arg Glu Thr Phe	Pro Gly Phe Glu	
2495		2500	2505	
Val Ser	Glu Leu Phe Glu	Thr Gln Ser Ile Asp	Thr Leu Leu Gly	
2510		2515	2520	
Ser Leu	Glu Gln Ala Pro	Leu Leu Ala Thr Leu	Thr Ala Pro Pro	
2525		2530	2535	
Gln Gln	Asp Met Leu Gln	Gln Leu Lys Gln Leu	Leu Ala Arg Thr	
2540		2545	2550	
Leu Lys	Leu Asp Ile Thr	Gln Ile Asp Thr Ser	Lys Thr Leu Glu	
2555		2560	2565	
Ser Tyr	Gly Val Asp Ser	Ile Val Ile Ile Glu	Leu Ala Asn Ala	
2570		2575	2580	
Leu Arg	Glu Arg Tyr Pro	Ser Leu Asp Ala Ser	Gln Leu Met Glu	
2585		2590	2595	
Thr Leu	Ser Ile Asp Arg	Leu Val Ala Gln Trp	Gln Ala Thr Glu	
2600		2605	2610	
Pro Ala	Val Pro Ala Glu	Pro Thr Ala Glu Pro	Pro Val Ala Asp	
2615		2620	2625	
Glu Asp	Ala Ala Ala Ile	Ile Gly Leu Ala Gly	Arg Phe Pro Gly	
2630		2635	2640	
Ala Asp	Thr Leu Glu Glu	Phe Trp Asn Asn Leu	Arg Asn Gly Gln	
2645		2650	2655	

Ser	Ser	Met	Gly	Glu	Val	Pro	Gly	Glu	Arg	Trp	Asp	His	Gln	His
2660						2665					2670			
Tyr	Phe	Asp	Ser	Glu	Arg	Gln	Ala	Pro	Gly	Lys	Thr	Tyr	Ser	Arg
2675						2680					2685			
Trp	Gly	Ala	Phe	Leu	Arg	Asp	Ile	Asp	Gly	Phe	Asp	Ala	Ala	Phe
2690						2695					2700			
Phe	Glu	Trp	Pro	Asp	Ser	Val	Ala	Leu	Glu	Ser	Asp	Pro	Gln	Ala
2705						2710					2715			
Arg	Ile	Phe	Leu	Glu	Gln	Ala	Tyr	Ala	Gly	Ile	Glu	Asp	Ala	Gly
2720						2725					2730			
Tyr	Thr	Pro	Gly	Ser	Leu	Ser	Lys	Ser	Gln	Arg	Val	Gly	Val	Phe
2735						2740					2745			
Val	Gly	Val	Met	Asn	Gly	Tyr	Tyr	Ser	Gly	Gly	Ala	Arg	Phe	Trp
2750						2755					2760			
Gln	Ile	Ala	Asn	Arg	Val	Ser	Tyr	Gln	Phe	Asp	Phe	Arg	Gly	Pro
2765						2770					2775			
Ser	Leu	Ala	Val	Asp	Thr	Ala	Cys	Ser	Ala	Ser	Leu	Thr	Ala	Ile
2780						2785					2790			
His	Leu	Ala	Leu	Glu	Ser	Leu	Arg	Ser	Gly	Ser	Cys	Glu	Val	Ala
2795						2800					2805			
Leu	Ala	Gly	Gly	Val	Asn	Leu	Leu	Val	Asp	Pro	Gln	Gln	Tyr	Leu
2810						2815					2820			
Asn	Leu	Ala	Gly	Ala	Ala	Met	Leu	Ser	Ala	Gly	Ala	Ser	Cys	Arg
2825						2830					2835			
Pro	Phe	Gly	Glu	Ala	Ala	Asp	Gly	Phe	Val	Ala	Gly	Glu	Ala	Cys
2840						2845					2850			
Gly	Val	Val	Leu	Leu	Lys	Pro	Leu	Lys	Gln	Ala	Arg	Ala	Asp	Gly
2855						2860					2865			
Asp	Val	Ile	His	Ala	Val	Ile	Arg	Gly	Ser	Met	Ile	Asn	Ala	Gly
2870						2875					2880			
Gly	His	Thr	Ser	Ala	Phe	Ser	Ser	Pro	Asn	Pro	Ala	Ala	Gln	Ala
2885						2890					2895			

Glu Val	Val Arg	Gln Ala	Leu	Gln Arg	Ala Gly	Val	Ala Pro	Asp
2900			2905			2910		
Ser Ile	Ser Tyr	Ile Glu	Ala	His Gly	Thr Gly	Thr	Val Leu	Gly
2915			2920			2925		
Asp Ala	Val Glu	Leu Gly	Ala	Leu Asn	Lys Val	Phe	Asp Lys	Arg
2930			2935			2940		
Ala Ala	Pro Cys	Pro Ile	Gly	Ser Leu	Lys Ala	Asn	Ile Gly	His
2945			2950			2955		
Ala Glu	Ser Ala	Ala Gly	Ile	Ala Gly	Leu Ala	Lys	Leu Val	Leu
2960			2965			2970		
Gln Phe	Arg His	Gly Glu	Leu	Val Pro	Ser Leu	Asn	Ala Phe	Pro
2975			2980			2985		
Leu Asn	Pro Tyr	Ile Glu	Phe	Gly Arg	Phe Gln	Val	Gln Gln	Gln
2990			2995			3000		
Pro Ala	Pro Trp	Pro Arg	Arg	Gly Ala	Gln Pro	Arg	Arg Ala	Gly
3005			3010			3015		
Leu Ser	Ala Phe	Gly Ala	Gly	Gly Ser	Asn Ala	His	Leu Val	Val
3020			3025			3030		
Glu Glu	Ala Pro	Ala Met	Ala	Pro Gly	Val Ser	Ile	Ser Ala	Ser
3035			3040			3045		
Ser Pro	Ala Leu	Ile Val	Leu	Ser Ala	Arg Thr	Leu	Pro Ala	Leu
3050			3055			3060		
Gln Gln	Arg Ala	Arg Asp	Leu	Leu Val	Trp Met	Gln	Ala Arg	Gln
3065			3070			3075		
Val Asp	Asp Val	Met Leu	Ala	Asp Val	Ala Tyr	Thr	Leu His	Leu
3080			3085			3090		
Gly Arg	Val Ala	Met Glu	Gln	Arg Leu	Ala Phe	Thr	Ala Gly	Ser
3095			3100			3105		
Ala Ala	Glu Leu	Ser Glu	Lys	Leu Gln	Ala Tyr	Leu	Gly His	Ala
3110			3115			3120		
Ile Arg	Ala Asp	Ile Tyr	Leu	Ser Glu	Asp Thr	Pro	Gly Lys	Pro
3125			3130			3135		

Ala Gly	Ala Pro Ile Val	Ala Glu Glu Asp Leu Leu	Thr Leu Met
3140		3145	3150
Asp Ala	Trp Ile Glu Lys Gly	Gln Tyr Gly Arg Leu	Leu Glu Tyr
3155		3160	3165
Trp Thr	Lys Gly Gln Pro Ile	Asp Trp Asn Lys Leu	Tyr Trp Arg
3170		3175	3180
Lys Leu	Tyr Ala Asp Gly Arg	Pro Arg Arg Ile Ser	Leu Pro Thr
3185		3190	3195
Tyr Pro	Phe Glu His Arg Arg	Tyr Trp Gln Thr Pro	Val Pro Gly
3200		3205	3210
Glu Arg	Ser Leu His Ala Thr	Ala Pro Ala Thr Arg	Glu Thr Val
3215		3220	3225
Ala Val	Gly Ala Met Pro Asp	Pro Ala Gly Ala Thr	Val Gln Ala
3230		3235	3240
Arg Leu	Cys Ala Leu Cys Gln	Val Leu Leu Gly Lys	Pro Val Thr
3245		3250	3255
Ala Gln	Met Asp Phe Phe Ala	Val Gly Gly His Ser	Val Leu Ala
3260		3265	3270
Ile Gln	Leu Val Ser Arg Ile	Arg Lys Ser Phe Gly	Val Glu Tyr
3275		3280	3285
Pro Val	Ser Ala Leu Phe Glu	Ser Ala Leu Leu Ser	Asp Met Ala
3290		3295	3300
Arg Gln	Ile Glu Gln Leu Arg	Val Asn Gly Val Ala	Lys Arg Met
3305		3310	3315
Pro Ala	Leu Leu Pro Ala Gly	Arg Val Gly Ala Ile	Pro Ala Thr
3320		3325	3330
Tyr Ala	Gln Glu Arg Leu Trp	Leu Val His Glu His	Met Ser Glu
3335		3340	3345
Gln Arg	Ser Ser Tyr Asn Ile	Thr Phe Ala Met His	Phe Arg Gly
3350		3355	3360
Val Asp	Phe Arg Ala Glu Ala	Met Arg Ala Ala Leu	Asn Ala Leu
3365		3370	3375

Val Val	Arg His Glu Val	Leu	Arg Thr Arg Phe	Leu	Ser Glu Asp
3380		3385		3390	
Gly Gln	Leu Gln Gln Val	Ile	Ala Ala Ser Leu Thr	Leu Glu Val	
3395		3400		3405	
Pro Val	Arg Glu Met Ser	Val	Glu Glu Val Asp	Leu	Leu Leu Ala
3410		3415		3420	
Ala Ser	Thr Arg Glu Thr	Phe	Asp Leu Arg Gln Gly	Pro Leu Phe	
3425		3430		3435	
Lys Ala	Arg Ile Leu Arg	Val	Ala Ala Asp His His	Val Val Leu	
3440		3445		3450	
Ser Ser	Ile His His Ile	Ile	Ser Asp Gly Trp Ser	Leu Gly Val	
3455		3460		3465	
Phe Asn	Arg Asp Leu His	Gln	Leu Tyr Glu Ala Cys	Leu Arg Gly	
3470		3475		3480	
Thr Pro	Pro Thr Leu Pro	Thr	Leu Ala Val Gln Tyr	Ala Asp Tyr	
3485		3490		3495	
Ala Leu	Trp Gln Arg Gln Trp	Glu Leu Ala Ala	Pro	Leu Ser Tyr	
3500		3505		3510	
Trp Thr	Arg Ala Leu Glu Gly	Tyr Asp Asp Gly	Leu	Asp Leu Pro	
3515		3520		3525	
Tyr Asp	Arg Pro Arg Gly Ala	Thr Arg Ala Trp Arg	Ala Gly Leu		
3530		3535		3540	
Val Lys	His Arg Tyr Pro	Pro	Gln Leu Ala Gln Gln	Leu Ala Ala	
3545		3550		3555	
Tyr Ser	Gln Gln Tyr Gln Ala	Thr Leu Phe Met	Ser	Leu Leu Ala	
3560		3565		3570	
Gly Leu	Ala Leu Val Leu Gly	Arg Tyr Ala Asp	Arg	Lys Asp Val	
3575		3580		3585	
Cys Ile	Gly Ala Thr Val	Ser	Gly Arg Asp Gln	Leu	Glu Leu Glu
3590		3595		3600	
Glu Leu	Ile Gly Phe Phe Ile	Asn Ile Leu Pro	Leu	Arg Val Asp	

3605		3610		3615
Leu Ser Gly Asp Pro Cys Leu Glu Glu Val Leu Leu Arg Thr Arg				
3620		3625		3630
Gln Val Val Leu Asp Gly Phe Ala His Gln Ser Val Pro Phe Glu				
3635		3640		3645
His Val Leu Gln Ala Leu Arg Arg Gln Arg Asp Ser Ser Gln Ile				
3650		3655		3660
Pro Leu Val Pro Val Met Leu Arg His Gln Asn Phe Pro Thr Gln				
3665		3670		3675
Glu Ile Gly Asp Trp Pro Glu Gly Val Arg Leu Thr Gln Met Glu				
3680		3685		3690
Leu Gly Leu Asp Arg Ser Thr Pro Ser Glu Leu Asp Trp Gln Phe				
3695		3700		3705
Tyr Gly Asp Gly Ser Ser Leu Glu Leu Thr Leu Glu Tyr Ala Gln				
3710		3715		3720
Asp Leu Phe Asp Glu Ala Thr Val Arg Arg Met Ile Ala His His				
3725		3730		3735
Gln Gln Ala Leu Glu Ala Met Val Ser Arg Pro Gln Leu Arg Val				
3740		3745		3750
Gly Lys Trp Asp Met Leu Thr Ala Glu Glu Arg Arg Leu Phe Ala				
3755		3760		3765
Ala Leu Asn Ala Thr Gly Thr Pro Arg Glu Trp Pro Ser Leu Ala				
3770		3775		3780
Gln Gln Phe Glu Arg Gln Ala Gln Ala Thr Pro Gln Ala Ile Ala				
3785		3790		3795
Cys Val Ser Asp Gly Gln Ser Trp Ser Tyr Ala Gln Leu Glu Ala				
3800		3805		3810
Arg Ala Asn Gln Leu Ala Gln Ala Leu Arg Gly Gln Gly Ala Gly				
3815		3820		3825
Arg Asp Val Arg Val Ala Val Gln Ser Ala Arg Thr Pro Glu Leu				
3830		3835		3840

Leu Met	Ala	Leu	Leu	Ala	Ile	Phe	Lys	Ala	Gly	Ala	Cys	Tyr	Val
3845					3850					3855			
Pro Ile	Asp	Pro	Ala	Tyr	Pro	Ala	Ala	Tyr	Arg	Glu	Gln	Ile	Leu
3860					3865					3870			
Ala Glu	Val	Gln	Val	Ser	Ile	Val	Leu	Glu	Gln	Asp	Glu	Leu	Ala
3875					3880					3885			
Leu Asp	Glu	Gln	Gly	Gln	Phe	His	Asn	Pro	Arg	Trp	Arg	Glu	Gln
3890					3895					3900			
Ala Pro	Thr	Pro	Leu	Gly	Leu	Arg	Glu	His	Pro	Gly	Asp	Leu	Ala
3905					3910					3915			
Cys Val	Met	Val	Thr	Ser	Gly	Ser	Thr	Gly	Arg	Pro	Lys	Gly	Val
3920					3925					3930			
Met Val	Pro	Tyr	Ala	Gln	Leu	His	Asn	Trp	Leu	His	Ala	Gly	Trp
3935					3940					3945			
Gln Arg	Ser	Ala	Phe	Glu	Ala	Gly	Glu	Arg	Val	Leu	Gln	Lys	Thr
3950					3955					3960			
Ser Ile	Ala	Phe	Ala	Val	Ser	Val	Lys	Glu	Leu	Leu	Ser	Gly	Leu
3965					3970					3975			
Leu Ala	Gly	Val	Glu	Gln	Val	Met	Leu	Pro	Asp	Glu	Gln	Val	Lys
3980					3985					3990			
Asp Ser	Leu	Ala	Leu	Ala	Arg	Ala	Ile	Glu	Gln	Trp	Gln	Val	Thr
3995					4000					4005			
Arg Leu	Tyr	Leu	Val	Pro	Ser	His	Leu	Gln	Ala	Leu	Leu	Asp	Ala
4010					4015					4020			
Thr Gln	Gly	Arg	Asp	Gly	Leu	Leu	His	Ser	Leu	Arg	His	Val	Val
4025					4030					4035			
Thr Ala	Gly	Glu	Ala	Leu	Pro	Ser	Ala	Val	Arg	Glu	Thr	Val	Arg
4040					4045					4050			
Val Arg	Leu	Pro	Gln	Val	Gln	Leu	Trp	Asn	Asn	Tyr	Gly	Cys	Thr
4055					4060					4065			
Glu Leu	Asn	Asp	Ala	Thr	Tyr	His	Arg	Ser	Asp	Thr	Val	Ala	Pro
4070					4075					4080			

Gly Thr Phe Val Pro Ile Gly Ala Pro Ile Ala Asn Thr Glu Val		
4085	4090	4095
Tyr Val Leu Asp Arg Gln Leu Arg Gln Val Pro Ile Gly Val Met		
4100	4105	4110
Gly Glu Leu His Val His Ser Val Gly Met Ala Arg Gly Tyr Trp		
4115	4120	4125
Asn Arg Pro Gly Leu Thr Ala Ser Arg Phe Ile Ala His Pro Tyr		
4130	4135	4140
Ser Glu Glu Pro Gly Thr Arg Leu Tyr Lys Thr Gly Asp Met Val		
4145	4150	4155
Arg Arg Leu Ala Asp Gly Thr Leu Glu Tyr Leu Gly Arg Gln Asp		
4160	4165	4170
Phe Glu Val Lys Val Arg Gly His Arg Val Asp Thr Arg Gln Val		
4175	4180	4185
Glu Ala Ala Leu Arg Ala Gln Pro Ala Val Ala Glu Ala Val Val		
4190	4195	4200
Ser Gly His Arg Val Asp Gly Asp Met Gln Leu Val Ala Tyr Val		
4205	4210	4215
Val Ala Arg Glu Gly Gln Ala Pro Ser Ala Gly Glu Leu Lys Gln		
4220	4225	4230
Gln Leu Ser Ala Gln Leu Pro Thr Tyr Met Leu Pro Thr Val Tyr		
4235	4240	4245
Gln Trp Leu Glu Gln Leu Pro Arg Leu Ser Asn Gly Lys Leu Asp		
4250	4255	4260
Arg Leu Ala Leu Pro Ala Pro Gln Val Val His Ala Gln Glu Tyr		
4265	4270	4275
Val Ala Pro Arg Asn Glu Ala Glu Gln Arg Leu Ala Ala Leu Phe		
4280	4285	4290
Ala Glu Val Leu Arg Val Glu Gln Val Gly Ile His Asp Asn Phe		
4295	4300	4305
Phe Ala Leu Gly Gly His Ser Leu Ser Ala Ser Gln Leu Ile Ser		
4310	4315	4320

Arg Ile	Arg Gln Ser Phe His	Val Asp Leu Pro Leu	Ser Arg Ile
4325	4330	4335	
Phe Glu	Ala Pro Thr Ile Glu	Gly Leu Val Arg Gln	Leu Ala Leu
4340	4345	4350	
Pro Ser	Glu Gly Gly Val Ala	Ser Ile Ala Arg Val	Ala Arg Asn
4355	4360	4365	
Arg Thr	Ile Pro Leu Ser Leu	Phe Gln Glu Arg Leu	Trp Phe Val
4370	4375	4380	
His Gln	His Met Pro Glu Gln	Arg Thr Ser Tyr Asn	Gly Thr Leu
4385	4390	4395	
Ala Leu	Arg Leu Arg Gly Pro	Leu Ser Val Glu Ala	Met Arg Ala
4400	4405	4410	
Ala Leu	Arg Ala Leu Val Leu	Arg His Glu Ile Leu	Arg Thr Arg
4415	4420	4425	
Phe Val	Leu Pro Thr Gly Ala	Ser Glu Pro Val Gln	Val Ile Asp
4430	4435	4440	
Glu His	Ser Asp Phe Gln Leu	Ser Val Gln Leu Val	Glu Asp Thr
4445	4450	4455	
Glu Ile	Ala Ser Leu Met Asp	Glu Leu Ala Ser His	Ile Tyr Asp
4460	4465	4470	
Leu Ala	Asn Gly Pro Leu Phe	Ile Ala Cys Leu Leu	Gln Leu Asp
4475	4480	4485	
Glu Gln	Glu His Val Leu Leu	Ile Gly Met His His	Leu Ile Tyr
4490	4495	4500	
Asp Ala	Trp Ser Gln Phe Thr	Val Met Asn Arg Asp	Leu Arg Val
4505	4510	4515	
Leu Tyr	His Arg His Leu Gly	Leu Ala Gly Gly Asp	Leu Pro Glu
4520	4525	4530	
Leu Pro	Ile Gln Tyr Ala Asp	Tyr Ala Ile Trp Gln	Arg Ala Gln
4535	4540	4545	
Asn Leu	Asp Ala Gln Leu Ala	Tyr Trp Gln Ala Met	Leu His Asp
4550	4555	4560	

Tyr Asp Asp Gly Leu Glu Leu Pro Tyr Asp Tyr Pro Arg Pro Arg
 4565 4570 4575

Asn Arg Thr Trp His Ala Ala Val Tyr Thr His Thr Tyr Pro Ala
 4580 4585 4590

Glu Leu Val Gln Arg Phe Ala Gly Phe Val Gln Ala His Gln Ser
 4595 4600 4605

Thr Leu Phe Ile Gly Leu Leu Ala Ser Phe Ala Val Val Leu Asn
 4610 4615 4620

Lys Tyr Thr Gly Arg Asp Asp Leu Cys Ile Gly Thr Thr Thr Ala
 4625 4630 4635

Gly Arg Thr His Leu Glu Leu Glu Asn Leu Ile Gly Phe Phe Ile
 4640 4645 4650

Asn Ile Leu Pro Leu Arg Leu Arg Leu Asp Gly Asp Pro Asp Val
 4655 4660 4665

Ala Glu Ile Met Arg Arg Thr Arg Leu Val Ala Met Ser Ala Phe
 4670 4675 4680

Glu Asn Gln Ala Leu Pro Phe Glu His Leu Leu Asn Ala Leu His
 4685 4690 4695

Lys Gln Arg Asp Thr Ser Arg Ile Pro Leu Val Pro Val Val Met
 4700 4705 4710

Arg His Gln Asn Phe Pro Asp Thr Ile Gly Asp Trp Ser Asp Gly
 4715 4720 4725

Ile Arg Thr Glu Val Ile Gln Arg Asp Leu Arg Ala Thr Pro Asn
 4730 4735 4740

Glu Met Asp Leu Gln Phe Phe Gly Asp Gly Thr Gly Leu Ser Val
 4745 4750 4755

Thr Val Glu Tyr Ala Ala Glu Leu Phe Ser Glu Ala Thr Ile Arg
 4760 4765 4770

Arg Leu Ile His His His Gln Leu Val Leu Glu Gln Met Leu Ala
 4775 4780 4785

Ala His Glu Ser Ala Thr Cys Pro Leu Asp Val Ala Asp
 4790 4795 4800

WO 02/024736

<210> 3
 <211> 14406
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(14406)
 <223>

<220>
 <221> misc_feature
 <222> (226)..(240)
 <223> Acyl-CoA ligase subdomain I

<220>
 <221> misc_feature
 <222> (486)..(493)
 <223> Acyl-CoA ligase subdomain II

<220>
 <221> misc_feature
 <222> (897)..(913)
 <223> Beta-ketoacyl synthase 1 subdomain I

<220>
 <221> misc_feature
 <222> (1038)..(1047)
 <223> Beta-ketoacyl synthase 1 subdomain II

<220>
 <221> misc_feature
 <222> (1080)..(1089)
 <223> Beta-ketoacyl synthase 1 subdomain III

<220>
 <221> misc_feature
 <222> (2777)..(2793)
 <223> Beta-ketoacyl synthase 2 subdomain I

<220>
 <221> misc_feature
 <222> (2918)..(2927)
 <223> Beta-ketoacyl synthase 2 subdomain II

<220>
 <221> misc_feature
 <222> (2955)..(2964)
 <223> Beta-ketoacyl synthase 2 subdomain III

<220>
 <221> misc_feature
 <222> (1812)..(1842)

<223> Beta-ketoacyl reductase domain

<220>

<221> misc_feature

<222> (667)..(678)

<223> Acyl carrier protein 1 domain

<220>

<221> misc_feature

<222> (2484)..(2495)

<223> Acyl carrier protein 2 domain

<220>

<221> misc_feature

<222> (2568)..(2579)

<223> Acyl carrier protein 3 domain

<220>

<221> misc_feature

<222> (3806)..(3811)

<223> Adenylation subdomain I

<220>

<221> misc_feature

<222> (3851)..(3861)

<223> Adenylation subdomain II

<220>

<221> misc_feature

<222> (3917)..(3922)

<223> Adenylation subdomain III

<220>

<221> misc_feature

<222> (3967)..(3970)

<223> Adenylation subdomain IV

<220>

<221> misc_feature

<222> (4063)..(4069)

<223> Adenylation subdomain V

<220>

<221> misc_feature

<222> (4114)..(4128)

<223> Adenylation subdomain VI

<220>

<221> misc_feature

<222> (4152)..(4157)

<223> Adenylation subdomain VII

<220>
<221> misc_feature
<222> (4170)..(4189)
<223> Adenylation subdomain VIII

<220>
<221> misc_feature
<222> (4239)..(4245)
<223> Adenylation subdomain IX

<220>
<221> misc_feature
<222> (4259)..(4264)
<223> Adenylation subdomain X

<220>
<221> misc_feature
<222> (3261)..(3271)
<223> Peptidyl carrier protein 1 domain

<220>
<221> misc_feature
<222> (4306)..(4316)
<223> Peptidyl carrier protein 2 domain

<220>
<221> misc_feature
<222> (3333)..(3342)
<223> Condensation domain 1 subdomain I

<220>
<221> misc_feature
<222> (3381)..(3389)
<223> Condensation domain 1 subdomain II

<220>
<221> misc_feature
<222> (3456)..(3465)
<223> Condensation domain 1 subdomain III

<220>
<221> misc_feature
<222> (3495)..(3501)
<223> Condensation domain 1 subdomain IV

<220>
<221> misc_feature
<222> (3606)..(3617)
<223> Condensation domain 1 subdomain V

<220>
<221> misc_feature
<222> (3641)..(3647)

<223> Condensation domain 1 subdomain VI

<220>

<221> misc_feature

<222> (3658)..(3665)

<223> Condensation domain 1 subdomain VII

<220>

<221> misc_feature

<222> (4374)..(4383)

<223> Condensation domain 2 subdomain I

<220>

<221> misc_feature

<222> (4421)..(4429)

<223> Condensation domain 2 subdomain II

<220>

<221> misc_feature

<222> (4498)..(4507)

<223> Condensation domain 2 subdomain III

<220>

<221> misc_feature

<222> (4538)..(4544)

<223> Condensation domain 2 subdomain IV

<220>

<221> misc_feature

<222> (4649)..(4659)

<223> Condensation domain 2 subdomain V

<220>

<221> misc_feature

<222> (4685)..(4691)

<223> Condensation domain 2 subdomain VI

<220>

<221> misc_feature

<222> (4701)..(4708)

<223> Condensation domain 2 subdomain VII

<400> 3

ttg ccc aat gcg ctc atg cag ata act ctt gta gcc gtc cag ttt gca 48
Leu Pro Asn Ala Leu Met Gln Ile Thr Leu Val Ala Val Gln Phe Ala 15
1 5 10

ggc gta ttg tta ggc gtc acc gct cgc gcg gcg atc ccc aat aag gcg 96
Gly Val Leu Leu Gly Val Thr Ala Arg Ala Ala Ile Pro Asn Lys Ala 30
20 25

ggt atg aga cgc gca tgg ccg ccc ttc ccg cag gcg tgc tgt cgc tct 144
Gly Met Arg Arg Ala Trp Pro Pro Phe Pro Gln Ala Cys Cys Arg Ser 45
35 40

att gct tac ctc atg cag aga tcg cca atg tcg ccg tta cag caa acg Ile Ala Tyr Leu Met Gln Arg Ser Pro Met Ser Pro Leu Gln Gln Thr 50 55 60	192
ctg cta acc cgc ctc gcc agt gcg gcc gcc tcc cgg aca atg atc gag Leu Leu Thr Arg Leu Ala Ser Ala Ala Ser Arg Thr Met Ile Glu 65 70 75 80	240
ttt ccg cgt ccg gag cac gca tcg cca caa tgt tgc gac gat gcc gag Phe Pro Arg Pro Glu His Ala Ser Pro Gln Cys Cys Asp Asp Ala Glu 85 90 95	288
ctt gcg cga ctg atc gtg cag ttg tcg gcg gga ctg caa ccg ctg gcg Leu Ala Arg Leu Ile Val Gln Leu Ser Ala Gly Leu Gln Pro Leu Ala 100 105 110	336
atg ccg ggt acc tac gtg atc att gcc gcg cca cat ggt ggt ttg ttc Met Pro Gly Thr Tyr Val Ile Ile Ala Ala Pro His Gly Gly Leu Phe 115 120 125	384
gcg gca gcc ctg ctt gcc tgt ttg cat gcc aac ctg gtg gcg gtg ccg Ala Ala Ala Leu Leu Ala Cys Leu His Ala Asn Leu Val Ala Val Pro 130 135 140	432
ttt cca ctg gat gtt gct cag cca aat gag ccg gaa cag gcc agg ctg Phe Pro Leu Asp Val Ala Gln Pro Asn Glu Arg Glu Gln Ala Arg Leu 145 150 155 160	480
gag acg atc cac gca caa ttg atg gag cat gcc aat gta gcg gtt ctg Glu Thr Ile His Ala Gln Leu Met Glu His Gly Asn Val Ala Val Leu 165 170 175	528
ctt gac gat gtc gcc gat cgc agt gcc ttc gcg cgc atg gcg cat gct Leu Asp Asp Val Ala Asp Arg Ser Ala Phe Ala Arg Met Ala His Ala 180 185 190	576
gcg ggc acc ttc ctg gcg acc ttc gcc gat cta aag cgc gaa tcg acc Ala Gly Thr Phe Leu Ala Thr Phe Ala Asp Leu Lys Arg Glu Ser Thr 195 200 205	624
agc gcc tcc ttg tgc ccg gcg tcg cct tcg gac gcc gcc ttg ctg ttg Ser Ala Ser Leu Cys Pro Ala Ser Pro Ser Asp Ala Ala Leu Leu Leu 210 215 220	672
ttt acc tct ggt tcc tcg ggt gag tcc aag ggc atc ctg ctt agc cac Phe Thr Ser Gly Ser Ser Gly Glu Ser Lys Gly Ile Leu Leu Ser His 225 230 235 240	720
cgc aac ctg cat cat cag atc cag gct ggc atc cgg cag tgg agc ttg Arg Asn Leu His His Gln Ile Gln Ala Gly Ile Arg Gln Trp Ser Leu 245 250 255	768
gac gag cat agc cat gtg gtg acc tgg ctt tct ccc gcg cac aac ttc Asp Glu His Ser His Val Val Thr Trp Leu Ser Pro Ala His Asn Phe 260 265 270	816
ggc ctg cat ttc ggc ttg ctg gca ccc tgg ttc agt ggc gcg acg gtc Gly Leu His Phe Gly Leu Leu Ala Pro Trp Phe Ser Gly Ala Thr Val 275 280 285	864
agt ttc atc cat ccg cac agt tat atg aaa cga ccc ggc ttc tgg ctg Ser Phe Ile His Pro His Ser Tyr Met Lys Arg Pro Gly Phe Trp Leu	912

290	295	300	
gag acg gtt gcg gct aga gac gcc acg cac atg gcc gcg ccg aac ttc			960
Glu Thr Val Ala Ala Arg Asp Ala Thr His Met Ala Ala Pro Asn Phe			
305	310	315	320
gcg ttc gac tac tgc tgc gac tgg gtg atg gtc gag cag ctt ccg ccg			1008
Ala Phe Asp Tyr Cys Cys Asp Trp Val Met Val Glu Gln Leu Pro Pro			
325	330		335
tct gcg ttg tct acg ctt acg cat atc gtg tgt ggc ggc gag ccg gtg			1056
Ser Ala Leu Ser Thr Leu Thr His Ile Val Cys Gly Gly Glu Pro Val			
340	345		350
cgc gcc tcg acc atg cag cgc ttc ttc gag aaa ttc gcc gga ctc ggt			1104
Arg Ala Ser Thr Met Gln Arg Phe Phe Glu Lys Phe Ala Gly Leu Gly			
355	360		365
gcg cgt acg cag act ttc atg ccg cac ttc ggc ttg tct gaa acc ggt			1152
Ala Arg Thr Gln Thr Phe Met Pro His Phe Gly Leu Ser Glu Thr Gly			
370	375		380
gcg ctg agt acc ttg gac gag gcg ccc caa cag cgc gtc ttg gaa cta			1200
Ala Leu Ser Thr Leu Asp Glu Ala Pro Gln Gln Arg Val Leu Glu Leu			
385	390		395
gat gcc gac gcc ttg aac aaa cgc aag cgc gtg gcg gca ggg gcg agc			1248
Asp Ala Asp Ala Leu Asn Lys Arg Lys Arg Val Ala Ala Gly Ala Ser			
405	410		415
cag gcg cgt gtg aca gtg ctc aat tgc ggc gcc gtc gac caa gat gtg			1296
Gln Ala Arg Val Thr Val Leu Asn Cys Gly Ala Val Asp Gln Asp Val			
420	425		430
gag ttg cgt atc gtc tgt cct gaa ggc gag acg ttg tgc aga cca gat			1344
Glu Leu Arg Ile Val Cys Pro Glu Gly Glu Thr Leu Cys Arg Pro Asp			
435	440		445
gag atc ggc gaa ata tgg gta aag tcg cct gcg atc gcc cgt ggc tac			1392
Glu Ile Gly Glu Ile Trp Val Lys Ser Pro Ala Ile Ala Arg Gly Tyr			
450	455		460
ctg ttt gcg aag ccc gcc gat cag cga cag ttc aac tgc agc atc cgt			1440
Leu Phe Ala Lys Pro Ala Asp Gln Arg Gln Phe Asn Cys Ser Ile Arg			
465	470		475
cat acc gac gat agc ggt tac ttt cgt acc ggc gac ctg ggt ttc att			1488
His Thr Asp Asp Ser Gly Tyr Phe Arg Thr Gly Asp Leu Gly Phe Ile			
485	490		495
gcc gat ggc tgt ctg tat gtc acc gga agg gta aag gag gtg ctg atc			1536
Ala Asp Gly Cys Leu Tyr Val Thr Gly Arg Val Lys Glu Val Leu Ile			
500	505		510
ata cgc ggt aag aat cat tac ccc gca cat atc gaa gcc tcg atc gcc			1584
Ile Arg Gly Lys Asn His Tyr Pro Ala His Ile Glu Ala Ser Ile Ala			
515	520		525
gct acc gca tcg cct ggc gcg ctg atg ccg gtg gtg ttc agc atc gag			1632
Ala Thr Ala Ser Pro Gly Ala Leu Met Pro Val Val Phe Ser Ile Glu			
530	535		540
cgg cag gac gag gag cgc gta gct gcg gtg atc gcc gtc aat cac ccg			1680

Arg Gln Asp Glu Glu Arg Val Ala Ala Val Ile Ala Val Asn His Pro 545 550 555 560	
tgg acg ccg gca gca tgc gcc gcg cag gca cac aag atc cgg caa cag Trp Thr Pro Ala Ala Cys Ala Ala Gln Ala His Lys Ile Arg Gln Gln 565 570 575	1728
gta gcc gac cag cat gga gtc gcc ctg gcg gag cta gcc ttt gcc gaa Val Ala Asp Gln His Gly Val Ala Leu Ala Glu Leu Ala Phe Ala Glu 580 585 590	1776
cac cgg cac gtg ttc ggc acc tat ccg gcc aaa ctg aag cgg cgc cta His Arg His Val Phe Gly Thr Tyr Pro Gly Lys Leu Lys Arg Arg Leu 595 600 605	1824
gtc aag gaa gcc tat gtc aac ggc cag ctg ccg ttg tta tgg cat gag Val Lys Glu Ala Tyr Val Asn Gly Gln Leu Pro Leu Leu Trp His Glu 610 615 620	1872
ggc aag aac ccg gac gta cca gcg gcc gcc gcg gac gat cgg cag gcg Gly Lys Asn Arg Asp Val Pro Ala Ala Ala Ala Asp Asp Arg Gln Ala 625 630 635 640	1920
caa cac gtg gcg gac ctg tgt cgg aag gtc ttt ttg ccg gtg ttg ggt Gln His Val Ala Asp Leu Cys Arg Lys Val Phe Leu Pro Val Leu Gly 645 650 655	1968
gtc gcg ccg ccg cat gcc caa tgg ccg ctg tgc gaa ctg gcg ctg gat Val Ala Pro Pro His Ala Gln Trp Pro Leu Cys Glu Leu Ala Leu Asp 660 665 670	2016
tcg ctc caa tgc gtg cgt ctt gcc ggt gcc atc gaa gag tgc tac ggc Ser Leu Gln Cys Val Arg Leu Ala Gly Ala Ile Glu Glu Cys Tyr Gly 675 680 685	2064
gtg cct ttc gaa ccc acg ttg cta ttc aag ctt gag acg gtc ggg gca Val Pro Phe Glu Pro Thr Leu Leu Phe Lys Leu Glu Thr Val Gly Ala 690 695 700	2112
atc gcc gaa tat gtc ctg gcg cac gga cgt cag gcg ccc acg ccg acg Ile Ala Glu Tyr Val Leu Ala His Gly Arg Gln Ala Pro Thr Pro Thr 705 710 715 720	2160
cgt gcg ccg gtg gca agc aca aca tgc tca gag gaa ccg atc gcc att Arg Ala Pro Val Ala Ser Thr Thr Cys Ser Glu Glu Pro Ile Ala Ile 725 730 735	2208
gtg gcg atg cac tgt gag gtg ccc gga gcg ggc gag aac act gaa gca Val Ala Met His Cys Glu Val Pro Gly Ala Gly Glu Asn Thr Glu Ala 740 745 750	2256
ttg tgg tcg ttc ctg ccg agc gac gtc aac gcg atc cgg ccg atc gaa Leu Trp Ser Phe Leu Arg Ser Asp Val Asn Ala Ile Arg Pro Ile Glu 755 760 765	2304
tca acg cgc ccg gac tta tgg gca gcg atg cgc gcc tat ccc ggc ctc Ser Thr Arg Pro Asp Leu Trp Ala Ala Met Arg Ala Tyr Pro Gly Leu 770 775 780	2352
gcg ggc gaa cag ctg ccg cgc tat gcg ggt ttc ctc gac gac gtt gat Ala Gly Glu Gln Leu Pro Arg Tyr Ala Gly Phe Leu Asp Asp Val Asp 785 790 795 800	2400
gct ttc gat gct gcg ttt ttc ggt atc tcg cgt cgc gag gcc gaa tgc	2448

Ala Phe Asp Ala Ala Phe Phe Gly Ile Ser Arg Arg Glu Ala Glu Cys	
805 810 815	
atg gac ccg cag cag cgc aaa gtg ctg gag atg gtg tgg aag ctg atc	2496
Met Asp Pro Gln Gln Arg Lys Val Leu Glu Met Val Trp Lys Leu Ile	
820 825 830	
gag caa gcc ggt cac gat ccg ctg tcc tgg ggc ggc cag ccg gtc ggc	2544
Glu Gln Ala Gly His Asp Pro Leu Ser Trp Gly Gly Gln Pro Val Gly	
835 840 845	
ctg ttc gtg ggt gcg cat acg tcc gac tat ggc gag ctg ctg gcg agc	2592
Leu Phe Val Gly Ala His Thr Ser Asp Tyr Gly Glu Leu Leu Ala Ser	
850 855 860	
cag ccg caa ctg atg gcc caa tgt ggc gct tac atc gat tcc ggt tcc	2640
Gln Pro Gln Leu Met Ala Gln Cys Gly Ala Tyr Ile Asp Ser Gly Ser	
865 870 875 880	
cat ttg acc atg att ccg aac cgg gct tcc cgc tgg ttc aat ttc acc	2688
His Leu Thr Met Ile Pro Asn Arg Ala Ser Arg Trp Phe Asn Phe Thr	
885 890 895	
ggc ccc agc gaa gta atc aac agc gct tgc tcc agc tcc ctg gtg gcg	2736
Gly Pro Ser Glu Val Ile Asn Ser Ala Cys Ser Ser Ser Leu Val Ala	
900 905 910	
ctg cat cgg gcg gtt caa tcc ctg cgc caa ggc gaa agc agt gtc gcc	2784
Leu His Arg Ala Val Gln Ser Leu Arg Gln Gly Glu Ser Ser Val Ala	
915 920 925	
ctg gta ctc ggc gtg aac ctt atc ctg gct ccc aag gtg ctg tta gcc	2832
Leu Val Leu Gly Val Asn Leu Ile Leu Ala Pro Lys Val Leu Leu Ala	
930 935 940	
agt gca agc gcg ggc atg ctt tcc ccc gat ggc cgc tgc aag acg ctt	2880
Ser Ala Ser Ala Gly Met Leu Ser Pro Asp Gly Arg Cys Lys Thr Leu	
945 950 955 960	
gac gcc gcc gcc gat ggc ttc gtg cgt tcc gaa ggg atc gca ggg gtg	2928
Asp Ala Ala Ala Asp Gly Phe Val Arg Ser Glu Gly Ile Ala Gly Val	
965 970 975	
ata ttg aag cca ctg gcg cag gcg ctg gcc gat ggt gac agg gtc tac	2976
Ile Leu Lys Pro Leu Ala Gln Ala Leu Ala Asp Gly Asp Arg Val Tyr	
980 985 990	
ggc cta gtc cgc ggc gtg gcg gtc aac cat ggc ggc cgt tcc aat tcc	3024
Gly Leu Val Arg Gly Val Ala Val Asn His Gly Gly Arg Ser Asn Ser	
995 1000 1005	
ttg cgt gct ccc aac gtc aac gcg cag cgg caa ctg ctg atc cgg	3069
Leu Arg Ala Pro Asn Val Asn Ala Gln Arg Gln Leu Leu Ile Arg	
1010 1015 1020	
act tac cag gaa gcc ggt gtc gag ccg gcc agc gtc ggt tat gtt	3114
Thr Tyr Gln Glu Ala Gly Val Glu Pro Ala Ser Val Gly Tyr Val	
1025 1030 1035	
gaa cta cac ggc act ggt acc agc ctg ggt gat ccg atc gaa atc	3159
Glu Leu His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Ile	
1040 1045 1050	

cag gcg	ctg aag	gaa gct	ttc att	gcg ttg	ggg gca	cag gcc	gcc	3204
Gln Ala	Leu Lys	Glu Ala	Phe Ile	Ala Leu	Gly Ala	Gln Ala	Ala	
1055			1060		1065			
ccg tca	aac tgc	ggc atc	ggc	tcg gtg	aag tcc	gcg	ctg ggc	3249
Pro Ser	Asn Cys	Gly Ile	Gly Ser	Val Lys	Ser Ala	Leu Gly	His	
1070			1075		1080			
cta gaa	gcc gct	gca ggc	ctg acc	ggc ctg	atc aag	gtg ctg	ctg	3294
Leu Glu	Ala Ala	Ala Gly	Leu Thr	Gly Leu	Ile Lys	Val Leu	Leu	
1085			1090		1095			
atg ctc	aag cac	ggc gag	cag gcc	ggc acg	cgc cat	ttc agc	acg	3339
Met Leu	Lys His	Gly Glu	Gln Ala	Gly Thr	Arg His	Phe Ser	Thr	
1100			1105		1110			
ctc aat	ccg ctg	atc gat	ttg cga	ggc acg	tca ttc	gaa gtg	gtg	3384
Leu Asn	Pro Leu	Ile Asp	Leu Arg	Gly Thr	Ser Phe	Glu Val	Val	
1115			1120		1125			
gcg cag	cat cgc	gca tgg	ccg	tcg cag	gtc ggc	att cac	ggc aca	3429
Ala Gln	His Arg	Ala Trp	Pro Ser	Gln Val	Gly Ile	His Gly	Thr	
1130			1135		1140			
ctc ttg	ccg cgt	cgc gcg	ggc	atc agc	tca ttc	ggc	ttc ggc	3474
Leu Leu	Pro Arg	Arg Ala	Gly Ile	Ser Ser	Phe Gly	Phe Gly	Gly	
1145			1150		1155			
gcc aat	gcg cat	gcg atc	gtg gaa	gag cat	gtc att	gcc acg	ccc	3519
Ala Asn	Ala His	Ala Ile	Val Glu	Glu His	Val Ile	Ala Thr	Pro	
1160			1165		1170			
ccc tcg	acg agc	tcc gct	ggc	ggc ccg	gta ggt	atc	gtg ttg	3564
Pro Ser	Thr Ser	Ser Ala	Gly Gly	Pro Val	Gly Ile	Val Leu	Ser	
1175			1180		1185			
gcc ggt	agt gaa	gct gtc	ttg	cgg caa	caa gtg	ctg	gcc ttg	3609
Ala Gly	Ser Glu	Ala Val	Leu Arg	Gln Gln	Val Leu	Ala Leu	Ser	
1190			1195		1200			
gcc tgg	cta agg	cag caa	tcg	ccg aca	ccc gcg	caa	atg atc	3654
Ala Trp	Leu Arg	Gln Gln	Ser Pro	Thr Pro	Ala Gln	Met Ile	Asp	
1205			1210		1215			
gtc gcc	tac acc	tta cag	gta	gga cgc	gca gcc	ctg	tcg cac	3699
Val Ala	Tyr Thr	Leu Gln	Val Gly	Arg Ala	Ala Leu	Ser His	Arg	
1220			1225		1230			
ttg gct	ttt agc	gcg acg	gac	gcc gag	cag gca	ttg	gcg agg	3744
Leu Ala	Phe Ser	Ala Thr	Asp Ala	Glu Gln	Ala Leu	Ala Arg	Leu	
1235			1240		1245			
gag ggt	cgt ctg	gcg ggc	gtg	atg gat	gcc gag	gtc	cat cac	3789
Glu Gly	Arg Leu	Ala Gly	Val Met	Asp Ala	Glu Val	His His	Gly	
1250			1255		1260			
gtc gtg	gat gct	gcc gca	acg	gct ccc	gaa cat	ggg	cgg cag	3834
Val Val	Asp Ala	Ala Ala	Thr Ala	Pro Glu	His Gly	Arg Gln	Thr	
1265			1270		1275			
cgc gaa	ggc ctt	gcc ggt	ttg	ctg cga	gcc tgg	act	cag ggc	3879
Arg Glu	Gly Leu	Ala Gly	Leu Leu	Arg Ala	Trp Thr	Gln Gly	Val	
1280			1285		1290			

cgc gtc	gat tgg	tcg gcg	ctg tac	ggc ata	cag cga	ccg cag	cgc	3924
Arg Val	Asp Trp	Ser Ala	Leu Tyr	Gly Ile	Gln Arg	Pro Gln	Arg	
1295			1300		1305			
gtt agc	ctg cct	gtc tac	ccc ttc	gct agg	gaa cgc	tat tgg	ctg	3969
Val Ser	Leu Pro	Val Tyr	Pro Phe	Ala Arg	Glu Arg	Tyr Trp	Leu	
1310			1315		1320			
ccc ggc	cag gct	atg cat	gcc gct	gcg gac	gct cat	ccg atg	ctg	4014
Pro Gly	Gln Ala	Met His	Ala Ala	Ala Asp	Ala His	Pro Met	Leu	
1325			1330		1335			
cag ctg	ttg cat	gcc aat	gcc aaa	cta cat	cgc tac	gcc ttg	cgt	4059
Gln Leu	Leu His	Ala Asn	Ala Lys	Leu His	Arg Tyr	Ala Leu	Arg	
1340			1345		1350			
agg tcc	ggc tgc	gca agc	ttt ctt	gtt gat	cat tgc	gtg gat	ggc	4104
Arg Ser	Gly Cys	Ala Ser	Phe Leu	Val Asp	His Cys	Val Asp	Gly	
1355			1360		1365			
cga cag	gta cta	ccg gca	gcc gtg	caa ctg	gaa ttg	gtg cgc	gcc	4149
Arg Gln	Val Leu	Pro Ala	Ala Val	Gln Leu	Glu Leu	Val Arg	Ala	
1370			1375		1380			
gtg gcg	cag cgg	gtc atg	gcg cag	gat gag	ggc tgt	atc gaa	ctg	4194
Val Ala	Gln Arg	Val Met	Ala Gln	Asp Glu	Gly Cys	Ile Glu	Leu	
1385			1390		1395			
gcg cag	gtc gcc	ttt ttg	cat ccc	ctc atg	atg gag	gag act	gag	4239
Ala Gln	Val Ala	Phe Leu	His Pro	Leu Met	Met Glu	Glu Thr	Glu	
1400			1405		1410			
ctg gag	gtc gaa	atc gaa	ctg tcg	aag agc	gat caa	gat gag	ttc	4284
Leu Glu	Val Glu	Ile Glu	Leu Ser	Lys Ser	Asp Gln	Asp Glu	Phe	
1415			1420		1425			
gat ttc	caa ctt	cac gat	gct cac	cgc caa	cag gtc	ttt agc	cag	4329
Asp Phe	Gln Leu	His Asp	Ala His	Arg Gln	Gln Val	Phe Ser	Gln	
1430			1435		1440			
ggg cac	gta cgt	cgc cgg	gtc tat	acg gcg	aca ccg	cgc ttg	gat	4374
Gly His	Val Arg	Arg Arg	Val Tyr	Thr Ala	Thr Pro	Arg Leu	Asp	
1445			1450		1455			
tta gcc	cag ctg	caa aag	ctt tgt	gcc gag	cgc gtg	ttg tcc	ggc	4419
Leu Ala	Gln Leu	Gln Lys	Leu Cys	Ala Glu	Arg Val	Leu Ser	Gly	
1460			1465		1470			
gaa gac	tgt tat	gcg cac	ttc acc	gcc tgc	gga ttg	cag ctc	ggc	4464
Glu Asp	Cys Tyr	Ala His	Phe Thr	Ala Cys	Gly Leu	Gln Leu	Gly	
1475			1480		1485			
gac cgg	ctc aaa	tcc gtg	caa tcg	atc ggc	tgc gga	cgc aat	ggc	4509
Asp Arg	Leu Lys	Ser Val	Gln Ser	Ile Gly	Cys Gly	Arg Asn	Gly	
1490			1495		1500			
gag ggc	gag ccg	atc gca	ttg ggt	gtc ctg	cgc ctg	cca cca	tca	4554
Glu Gly	Glu Pro	Ile Ala	Leu Gly	Val Leu	Arg Leu	Pro Pro	Ser	
1505			1510		1515			
agc gtt	gaa gac	agc cat	gtg ctg	cct cct	agc ctg	ctt gat	ggc	4599
Ser Val	Glu Asp	Ser His	Val Leu	Pro Pro	Ser Leu	Leu Asp	Gly	
1520			1525		1530			

gcc ttg	cag tgt	agc ctt	ggc	ttg cag	cgt gat	gtc	gag cac	atc	4644
Ala Leu	Gln Cys	Ser Leu	Gly	Leu Gln	Arg Asp	Val	Glu His	Ile	
1535			1540			1545			
gcc atg	cca tac	acg ctg	gag	cgg atg	acg gtg	cat	gcg ccg	att	4689
Ala Met	Pro Tyr	Thr Leu	Glu	Arg Met	Thr Val	His	Ala Pro	Ile	
1550			1555			1560			
cct ccc	gag gcc	tgg gtg	ctg	ctg cgt	cac ggc	cat	gca gcc	aga	4734
Pro Pro	Glu Ala	Trp Val	Leu	Leu Arg	His Gly	His	Ala Ala	Arg	
1565			1570			1575			
cag tcc	ctg gac	atc gat	ctc	ctg gat	tcc gaa	ggc	agg gtc	tgc	4779
Gln Ser	Leu Asp	Ile Asp	Leu	Leu Asp	Ser Glu	Gly	Arg Val	Cys	
1580			1585			1590			
gtc agc	ctc ggc	aat tac	acc	ggc cgt	gca ccg	aaa	gcc gtt	tcc	4824
Val Ser	Leu Gly	Asn Tyr	Thr	Gly Arg	Ala Pro	Lys	Ala Val	Ser	
1595			1600			1605			
gcc gtc	agg gcg	ctt gtc	ttg	gca ccg	gtc tgg	caa	gcg ttg	acc	4869
Ala Val	Arg Ala	Leu Val	Leu	Ala Pro	Val Trp	Gln	Ala Leu	Thr	
1610			1615			1620			
gaa acg	gcg ccg	gca tgg	ccc	gat ccg	gcc gaa	cgc	atc gtt	acg	4914
Glu Thr	Ala Pro	Ala Trp	Pro	Asp Pro	Ala Glu	Arg	Ile Val	Thr	
1625			1630			1635			
gta gga	gac gat	gca tgg	cgt	agt cac	ttc ggt	ttc	gac gag	ccg	4959
Val Gly	Asp Asp	Ala Trp	Arg	Ser His	Phe Gly	Phe	Asp Glu	Pro	
1640			1645			1650			
gcc ttg	tcc ctg	gag gac	agc	gtc gaa	gtc atc	gcg	acg cga	ctg	5004
Ala Leu	Ser Leu	Glu Asp	Ser	Val Glu	Val Ile	Ala	Thr Arg	Leu	
1655			1660			1665			
ggc cag	agc ggc	aag ttc	gat	cat cta	gtc tgg	atc	gtg ccg	ata	5049
Gly Gln	Ser Gly	Lys Phe	Asp	His Leu	Val Trp	Ile	Val Pro	Ile	
1670			1675			1680			
gcc gag	agt gaa	acc gat	att	gca gcg	caa ggt	tca	gcg gcg	atc	5094
Ala Glu	Ser Glu	Thr Asp	Ile	Ala Ala	Gln Gly	Ser	Ala Ala	Ile	
1685			1690			1695			
gcc ggt	ttc cgg	ttg gtc	aag	gcg ttg	ctt gcg	ttg	ggc tat	gcg	5139
Ala Gly	Phe Arg	Leu Val	Lys	Ala Leu	Leu Ala	Leu	Gly Tyr	Ala	
1700			1705			1710			
cat cgc	ccg ctg	ggt ctc	acc	gtg ctg	act cgc	caa	gcc ctt	acg	5184
His Arg	Pro Leu	Gly Leu	Thr	Val Leu	Thr Arg	Gln	Ala Leu	Thr	
1715			1720			1725			
cgg cag	ccg tcg	cac gcg	gca	gtg cac	ggg ctg	atc	ggg acg	ctg	5229
Arg Gln	Pro Ser	His Ala	Ala	Val His	Gly Leu	Ile	Gly Thr	Leu	
1730			1735			1740			
gcc aag	gaa tac	tgc aac	tgg	aaa atc	cgt ctg	ctc	gac ctg	ccg	5274
Ala Lys	Glu Tyr	Cys Asn	Trp	Lys Ile	Arg Leu	Leu	Asp Leu	Pro	
1745			1750			1755			
agc gta	aaa tct	tgg ccg	caa	tgg gag	caa ttg	cgg	tcg ttg	cct	5319
Ser Val	Lys Ser	Trp Pro	Gln	Trp Glu	Gln Leu	Arg	Ser Leu	Pro	

1760	1765	1770	
tgg cat gcg cag ggc gaa gcc	ctg atc ggc cgt ggg	act tgt tgg	5364
Trp His Ala Gln Gly Glu Ala	Leu Ile Gly Arg Gly Thr	Cys Trp	
1775	1780	1785	
tat cgg cgg cag ttg tgt gaa	gtg ctg ccg ctg ccg	tcg ttg gaa	5409
Tyr Arg Arg Gln Leu Cys Glu	Val Leu Pro Leu Pro	Ser Leu Glu	
1790	1795	1800	
ccg ccg ccg tac cgc gta ggc	ggc gtc tac gtc gtg	atc ggc ggc	5454
Pro Pro Pro Tyr Arg Val Gly	Gly Val Tyr Val Val	Ile Gly Gly	
1805	1810	1815	
gct ggc ggc ttg ggt gaa gta	ttg agc gaa cac ttg	atc cgc acg	5499
Ala Gly Gly Leu Gly Glu Val	Leu Ser Glu His Leu	Ile Arg Thr	
1820	1825	1830	
tac gac gcg cag ctg atc tgg	atc ggc cgg cgc gtg	ctg gac gaa	5544
Tyr Asp Ala Gln Leu Ile Trp	Ile Gly Arg Arg Val	Leu Asp Glu	
1835	1840	1845	
ggc att gcg cgc aag cag acc	cgg ctt gcg tcg ctg	ggc cgc gca	5589
Gly Ile Ala Arg Lys Gln Thr	Arg Leu Ala Ser Leu	Gly Arg Ala	
1850	1855	1860	
ccg cat tac atc tcc gcg gac	gcg agt gac ccg gct	gcc ctg cag	5634
Pro His Tyr Ile Ser Ala Asp	Ala Ser Asp Pro Ala	Ala Leu Gln	
1865	1870	1875	
gcg gca cat aat gag atc gtt	gcg ctg cat ggc cag	ccc cat ggg	5679
Ala Ala His Asn Glu Ile Val	Ala Leu His Gly Gln	Pro His Gly	
1880	1885	1890	
ctc atc cta agc aac atc gtg	ctg aag gat gcc agt	ctg gct cgt	5724
Leu Ile Leu Ser Asn Ile Val	Leu Lys Asp Ala Ser	Leu Ala Arg	
1895	1900	1905	
atg gag gaa gcc gat ttc cgt	gac gtg ctg gcc gcg	aaa ctc gac	5769
Met Glu Glu Ala Asp Phe Arg	Asp Val Leu Ala Ala	Lys Leu Asp	
1910	1915	1920	
gtc agc gtg tgt gcg gca cag	gtg ttc ggc acg gcc	ccc ctt gat	5814
Val Ser Val Cys Ala Ala Gln	Val Phe Gly Thr Ala	Pro Leu Asp	
1925	1930	1935	
ttc gtg ctg ttt ttt tct tcc	atc cag agc act acc	aag gcg gcc	5859
Phe Val Leu Phe Phe Ser Ser	Ile Gln Ser Thr Thr	Lys Ala Ala	
1940	1945	1950	
ggg caa ggt aac tac gcc gcc	ggc tgc tgc tat gtc	gac gct ttc	5904
Gly Gln Gly Asn Tyr Ala Ala	Gly Cys Cys Tyr Val	Asp Ala Phe	
1955	1960	1965	
ggc gag cta tgg gcg cgc cgg	ggc ttg agg gta aag	acc atc aac	5949
Gly Glu Leu Trp Ala Arg Arg	Gly Leu Arg Val Lys	Thr Ile Asn	
1970	1975	1980	
tgg ggc tac tgg ggc agc gtg	ggc gtc gta gcg ggc	gag gac tat	5994
Trp Gly Tyr Trp Gly Ser Val	Gly Val Val Ala Gly	Glu Asp Tyr	
1985	1990	1995	
cgc cgg cgc atg gcg caa aaa	cac atg gct tcg att	gag ggt gcc	6039

Arg Arg	Arg Met	Ala Gln	Lys	His Met	Ala Ser	Ile	Glu Gly	Ala	
2000			2005			2010			
gaa gcg	atg cag	gtg ttg	tcg	cag ttg	ttg tgt	gcg	ccg ttg	caa	6084
Glu Ala	Met Gln	Val Leu	Ser	Gln Leu	Leu Cys	Ala	Pro Leu	Gln	
2015			2020			2025			
cgg ctt	gcc tac	gtc aag	atc	gac gat	gct aac	gca	atg cgc	gct	6129
Arg Leu	Ala Tyr	Val Lys	Ile	Asp Asp	Ala Asn	Ala	Met Arg	Ala	
2030			2035			2040			
ctg ggc	gta gta	gag gac	gag	agc gtg	caa atc	cct	gtg cac	gca	6174
Leu Gly	Val Val	Glu Asp	Glu	Ser Val	Gln Ile	Pro	Val His	Ala	
2045			2050			2055			
ccg gcc	gag cct	ccc aga	ggg	cag cct	ggg	ccc	gtg	gtc gag	6219
Pro Ala	Glu Pro	Pro Arg	Gly	Gln Pro	Gly Pro	Val	Val Glu	Leu	
2060			2065			2070			
tcg gtg	aat ctg	gat gcc	cgg	cgc gaa	cgg gaa	act	ttg ctg	gcg	6264
Ser Val	Asn Leu	Asp Ala	Arg	Arg Glu	Arg Glu	Thr	Leu Leu	Ala	
2075			2080			2085			
gcc tgg	ctg ctt	gag ttg	atc	gag caa	ctc ggt	ggg	ttt ccg	ccg	6309
Ala Trp	Leu Leu	Glu Leu	Ile	Glu Gln	Leu Gly	Gly	Phe Pro	Pro	
2090			2095			2100			
gca agt	ttc gac	atc gct	acg	ctt gcg	caa cgc	ctg	cac atc	gta	6354
Ala Ser	Phe Asp	Ile Ala	Thr	Leu Ala	Gln Arg	Leu	His Ile	Val	
2105			2110			2115			
ccc gcc	tat cga	agc tgg	ctg	gaa cac	agc gtg	cgg	atg ctc	ggc	6399
Pro Ala	Tyr Arg	Ser Trp	Leu	Glu His	Ser Val	Arg	Met Leu	Gly	
2120			2125			2130			
gtg tat	ggg tac	ctc aga	gcg	acg ggg	gaa agc	cga	ttc cag	ctg	6444
Val Tyr	Gly Tyr	Leu Arg	Ala	Thr Gly	Glu Ser	Arg	Phe Glu	Leu	
2135			2140			2145			
gcc gac	aag ccg	ccc gat	gat	gcc agg	ggg gcc	tgg	aac gcg	cat	6489
Ala Asp	Lys Pro	Pro Asp	Asp	Ala Arg	Gly Ala	Trp	Asn Ala	His	
2150			2155			2160			
gtg cac	gag gcc	agc gtc	gaa	gcc ggt	gaa gag	gca	cag cgg	cgt	6534
Val His	Glu Ala	Ser Val	Glu	Ala Gly	Glu Glu	Ala	Gln Arg	Arg	
2165			2170			2175			
ctg ctc	gat cgc	tgc atg	cgg	gcg ttg	ccg gcg	gtc	ctt cga	ggc	6579
Leu Leu	Asp Arg	Cys Met	Arg	Ala Leu	Pro Ala	Val	Leu Arg	Gly	
2180			2185			2190			
gaa cgc	aag gcc	acc gaa	ttg	ctg ttt	ccg gaa	ggg	tcg atg	gcg	6624
Glu Arg	Lys Ala	Thr Glu	Leu	Leu Phe	Pro Glu	Gly	Ser Met	Ala	
2195			2200			2205			
tgg gtc	gag ggt	atc tac	cag	aac aac	ccg ctt	gcc	gat tac	ttc	6669
Trp Val	Glu Gly	Ile Tyr	Gln	Asn Asn	Pro Leu	Ala	Asp Tyr	Phe	
2210			2215			2220			
aac gca	caa cta	gtc acg	cga	ctg att	gcc tac	ttg	aga cga	cga	6714
Asn Ala	Gln Leu	Val Thr	Arg	Leu Ile	Ala Tyr	Leu	Arg Arg	Arg	
2225			2230			2235			
cta gag	tcg acg	cct acg	gcg	cgc ctg	aag ctg	tgc	gag atc	ggc	6759

Leu Glu	Ser Thr	Pro Thr	Ala	Arg Leu Lys Leu Cys	Glu Ile Gly	
2240			2245		2250	
gcc ggc	agc ggt	ggt act	act	gca agc	gtg cta	caa cag ttg cag
Ala Gly	Ser Gly	Gly Thr	Thr	Ala Ser	Val Leu	Gln Gln Leu Gln
2255			2260			2265
gca tat	ggt gag	cat att	gag	gaa tat	ctc tat	acc gac ctg tcg
Ala Tyr	Gly Glu	His Ile	Glu	Glu Tyr	Leu Tyr	Thr Asp Leu Ser
2270			2275			2280
cct gtc	ttc ctg	cat cat	gcg	gaa aaa	cac tat	cag cca cga gcg
Pro Val	Phe Leu	His His	Ala	Glu Lys	His Tyr	Gln Pro Arg Ala
2285			2290			2295
cct tat	ttg agg	acc gcc	tgt	ttc gac	gta gcg	cgc gcg ccg acg
Pro Tyr	Leu Arg	Thr Ala	Cys	Phe Asp	Val Ala	Arg Ala Pro Thr
2300			2305			2310
gcg cag	gcc ctg	gaa tct	ggc	ggc tac	gac gtg	gtg att gcc gcc
Ala Gln	Ala Leu	Glu Ser	Gly	Gly Tyr	Asp Val	Val Ile Ala Ala
2315			2320			2325
aac gta	ctg cat	gct acg	cgc	gat atc	gcc aag	acc ttg cgc aat
Asn Val	Leu His	Ala Thr	Arg	Asp Ile	Ala Lys	Thr Leu Arg Asn
2330			2335			2340
gcg aag	gca ctc	ctc aaa	cct	ggc ggt	ctg ctc	ttg ctc aac gaa
Ala Lys	Ala Leu	Leu Lys	Pro	Gly Gly	Leu Leu	Leu Leu Asn Glu
2345			2350			2355
gtg atc	gag cgc	agc ctc	gtc	ttg cac	ctg act	ttc ggt ctg ctg
Val Ile	Glu Arg	Ser Leu	Val	Leu His	Leu Thr	Phe Gly Leu Leu
2360			2365			2370
gag agc	ttg tgg	ttg ccc	cag	gac aag	atc ttg	cgc ctt gcc ggc
Glu Ser	Trp Trp	Leu Pro	Gln	Asp Lys	Ile Leu	Arg Leu Ala Gly
2375			2380			2385
tcg ccg	ttg ctg	gct tgc	gcc	acc tgg	cgc agc	ctg ctg gag gct
Ser Pro	Leu Leu	Ala Cys	Ala	Thr Trp	Arg Ser	Leu Leu Glu Ala
2390			2395			2400
gag ggt	ttt gcg	ggg ctg	agc	gtg cac	agg gcg	caa ccc gat gcc
Glu Gly	Phe Ala	Gly Leu	Ser	Val His	Arg Ala	Gln Pro Asp Ala
2405			2410			2415
ggg cag	gcc atc	atc tgt	gcc	tac agc	gat ggg	ata gtg cgg caa
Gly Gln	Ala Ile	Ile Cys	Ala	Tyr Ser	Asp Gly	Ile Val Arg Gln
2420			2425			2430
gcc agt	acg atc	gag gtt	gcg	cgg aat	gaa aaa	gta acc gtt ccg
Ala Ser	Thr Ile	Glu Val	Ala	Arg Asn	Glu Lys	Val Thr Val Pro
2435			2440			2445
tcg cag	ccg gcg	gaa gcc	ggg	gaa tcg	ccg ctg	gat ctg gtc aaa
Ser Gln	Pro Ala	Glu Ala	Gly	Glu Ser	Pro Leu	Asp Leu Val Lys
2450			2455			2460
aaa ctg	ctt gga	cgc att	ctg	aaa atg	gat ccg	gcc aca ctc gat
Lys Leu	Leu Gly	Arg Ile	Leu	Lys Met	Asp Pro	Ala Thr Leu Asp
2465			2470			2475

acc agc	cac ccg ctg gag tac	tac ggt gtc gat tcg	atc gtg gcg	7479
Thr Ser	His Pro Leu Glu Tyr	Tyr Gly Val Asp Ser	Ile Val Ala	
2480	2485	2490		
atc gaa	ctg gct atg gca ctg	cgc gag aca ttc ccg	ggg ttt gaa	7524
Ile Glu	Leu Ala Met Ala Leu	Arg Glu Thr Phe Pro	Gly Phe Glu	
2495	2500	2505		
gtc agc	gag ctg ttt gaa acg	caa tcc atc gat acc	ttg ttg ggc	7569
Val Ser	Glu Leu Phe Glu Thr	Gln Ser Ile Asp Thr	Leu Leu Gly	
2510	2515	2520		
tct ctt	gag cag gct cct ctc	ctt gct acc ctc aca	gct ccg ccg	7614
Ser Leu	Glu Gln Ala Pro Leu	Leu Ala Thr Leu Thr	Ala Pro Pro	
2525	2530	2535		
caa caa	gac atg ctg cag cag	ctg aaa caa ctg ctg	gcg cgt acg	7659
Gln Gln	Asp Met Leu Gln Gln	Leu Lys Gln Leu Leu	Ala Arg Thr	
2540	2545	2550		
ctg aag	ctg gac att acg cag	atc gac acg agc aag	acg ctg gag	7704
Leu Lys	Leu Asp Ile Thr Gln	Ile Asp Thr Ser Lys	Thr Leu Glu	
2555	2560	2565		
agc tat	ggg gtc gac tcc atc	gtc atc atc gaa tta	gcc aac gcc	7749
Ser Tyr	Gly Val Asp Ser Ile	Val Ile Ile Glu Leu	Ala Asn Ala	
2570	2575	2580		
ttg cgt	gag cgc tat ccg agc	ttg gac gcg tca cag	ctg atg gaa	7794
Leu Arg	Glu Arg Tyr Pro Ser	Leu Asp Ala Ser Gln	Leu Met Glu	
2585	2590	2595		
acc tta	tcg atc gac cgg ctg	gtt gcc caa tgg cag	gca acg gag	7839
Thr Leu	Ser Ile Asp Arg Leu	Val Ala Gln Trp Gln	Ala Thr Glu	
2600	2605	2610		
ccc gcc	gta ccg gca gag cca	aca gcg gaa ccg ccg	gta gcc gac	7884
Pro Ala	Val Pro Ala Glu Pro	Thr Ala Glu Pro Pro	Val Ala Asp	
2615	2620	2625		
gaa gac	gcc gct gcc atc atc	gga ctg gcc ggc cgc	ttt cca ggc	7929
Glu Asp	Ala Ala Ala Ile Ile	Gly Leu Ala Gly Arg	Phe Pro Gly	
2630	2635	2640		
gcg gac	acg ttg gag gag ttc	tgg aac aac ctg cgc	aac ggc caa	7974
Ala Asp	Thr Leu Glu Glu Phe	Trp Asn Asn Leu Arg	Asn Gly Gln	
2645	2650	2655		
agc agt	atg gga gag gtg cca	ggc gag cgc tgg gat	cac cag cac	8019
Ser Ser	Met Gly Glu Val Pro	Gly Glu Arg Trp Asp	His Gln His	
2660	2665	2670		
tac ttc	gac agt gaa cgc cag	gca ccg ggc aag acg	tat agc cgc	8064
Tyr Phe	Asp Ser Glu Arg Gln	Ala Pro Gly Lys Thr	Tyr Ser Arg	
2675	2680	2685		
tgg ggt	gcg ttt ctg agg gac	ata gac ggc ttc gat	gca gcc ttc	8109
Trp Gly	Ala Phe Leu Arg Asp	Ile Asp Gly Phe Asp	Ala Ala Phe	
2690	2695	2700		
ttt gaa	tgg ccc gac agc gtc	gcg ctg gaa tcg gat	ccg caa gcg	8154
Phe Glu	Trp Pro Asp Ser Val	Ala Leu Glu Ser Asp	Pro Gln Ala	
2705	2710	2715		

cgg ata ttt cta gag cag gcc	tat gcc ggg atc gaa	gat gcc ggc	8199
Arg Ile Phe Leu Glu Gln Ala	Tyr Ala Gly Ile Glu	Asp Ala Gly	
2720	2725	2730	
tac acg cct ggc tcg ctc agc	aag agc caa cgc gta	ggg gta ttc	8244
Tyr Thr Pro Gly Ser Leu Ser	Lys Ser Gln Arg Val	Gly Val Phe	
2735	2740	2745	
gta ggt gtg atg aat ggt tac	tac agc ggc gga gcg	cgc ttc tgg	8289
Val Gly Val Met Asn Gly Tyr	Tyr Ser Gly Gly Ala	Arg Phe Trp	
2750	2755	2760	
caa atc gcc aac cgc gtg tcg	tac cag ttc gat ttt	cgc ggg cca	8334
Gln Ile Ala Asn Arg Val Ser	Tyr Gln Phe Asp Phe	Arg Gly Pro	
2765	2770	2775	
agc ctg gcg gtg gat acc gcc	tgt tcg gct tcg ctc	acc gcg atc	8379
Ser Leu Ala Val Asp Thr	Cys Ser Ala Ser Leu	Thr Ala Ile	
2780	2785	2790	
cac ctg gcg ctg gaa agc ctg	cgc agc ggc agt tgc	gag gtc gca	8424
His Leu Ala Leu Glu Ser Leu	Arg Ser Gly Ser Cys	Glu Val Ala	
2795	2800	2805	
ctg gcc ggt ggc gtg aat ctg	ctg gtc gat ccg cag	caa tat ctt	8469
Leu Ala Gly Gly Val Asn Leu	Leu Val Asp Pro Gln	Gln Tyr Leu	
2810	2815	2820	
aat ttg gct ggc gcc gcg atg	ctc tcc gcc ggc gcc	agc tgt cgg	8514
Asn Leu Ala Gly Ala Ala Met	Leu Ser Ala Gly Ala	Ser Cys Arg	
2825	2830	2835	
ccg ttc ggc gag gcc gcg gac	ggg ttc gtg gcc ggc	gaa gcc tgc	8559
Pro Phe Gly Glu Ala Ala Asp	Gly Phe Val Ala Gly	Glu Ala Cys	
2840	2845	2850	
ggc gtg gtg ctg ctc aag ccg	ctc aag caa gcg agg	gcc gat ggc	8604
Gly Val Val Leu Leu Lys Pro	Leu Lys Gln Ala Arg	Ala Asp Gly	
2855	2860	2865	
gat gtg atc cat gcc gta atc	agg ggc agc atg atc	aat gcc ggt	8649
Asp Val Ile His Ala Val Ile	Arg Gly Ser Met Ile	Asn Ala Gly	
2870	2875	2880	
ggg cac acc agc gcg ttc tcc	tcg cct aac cct gcc	gcc cag gcc	8694
Gly His Thr Ser Ala Phe Ser	Ser Pro Asn Pro Ala	Ala Gln Ala	
2885	2890	2895	
gaa gtc gtg cgg cag gcc ttg	cag cgc gcg ggc gtg	gcg ccc gat	8739
Glu Val Val Arg Gln Ala Leu	Gln Arg Ala Gly Val	Ala Pro Asp	
2900	2905	2910	
tcg atc agc tac atc gag gcg	cat ggc acc ggc acc	gta cta ggc	8784
Ser Ile Ser Tyr Ile Glu Ala	His Gly Thr Gly Thr	Val Leu Gly	
2915	2920	2925	
gat gca gtg gag ttg ggt gct	ttg aat aaa gtg ttc	gac aag cgc	8829
Asp Ala Val Glu Leu Gly Ala	Leu Asn Lys Val Phe	Asp Lys Arg	
2930	2935	2940	
gcg gcg cca tgc ccg atc ggc	tcg ctg aag gcg aac	atc ggc cat	8874
Ala Ala Pro Cys Pro Ile Gly	Ser Leu Lys Ala Asn	Ile Gly His	
2945	2950	2955	

gcc gaa agc gcc gcg ggc atc gcc ggc ctg gcc aag ctg gta ttg Ala Glu Ser Ala Ala Gly Ile Ala Gly Leu Ala Lys Leu Val Leu 2960 2965 2970	8919
cag ttc agg cat ggc gag ttg gtg cct agt ctg aat gcg ttt ccc Gln Phe Arg His Gly Glu Leu Val Pro Ser Leu Asn Ala Phe Pro 2975 2980 2985	8964
ttg aat ccc tat att gag ttc ggt cgc ttc cag gta caa cag cag Leu Asn Pro Tyr Ile Glu Phe Gly Arg Phe Gln Val Gln Gln Gln 2990 2995 3000	9009
ccg gca ccg tgg ccg cgc cgt ggc gcc cag ccg cgg cgc gcc ggg Pro Ala Pro Trp Pro Arg Arg Gly Ala Gln Pro Arg Arg Ala Gly 3005 3010 3015	9054
tta tct gcc ttc ggt gct ggc gga tgc aat gcg cac cta gtg gta Leu Ser Ala Phe Gly Ala Gly Gly Ser Asn Ala His Leu Val Val 3020 3025 3030	9099
gag gaa gct ccg gct atg gct ccc ggg gtc tgc atc agc gcc agc Glu Glu Ala Pro Ala Met Ala Pro Gly Val Ser Ile Ser Ala Ser 3035 3040 3045	9144
tct cca gcc ttg atc gtg ctt tgc gcg cga acg ctg cct gcc ttg Ser Pro Ala Leu Ile Val Leu Ser Ala Arg Thr Leu Pro Ala Leu 3050 3055 3060	9189
caa cag cgt gct cgc gat ctg ctc gtc tgg atg caa gcg cgg cag Gln Gln Arg Ala Arg Asp Leu Leu Val Trp Met Gln Ala Arg Gln 3065 3070 3075	9234
gtg gat gac gtc atg ctg gcc gac gtt gct tat acg ctg cac ttg Val Asp Asp Val Met Leu Ala Asp Val Ala Tyr Thr Leu His Leu 3080 3085 3090	9279
ggc cgc gtc gcg atg gag caa cgc ctg gct ttt acc gct ggc tgc Gly Arg Val Ala Met Glu Gln Arg Leu Ala Phe Thr Ala Gly Ser 3095 3100 3105	9324
gct gcc gag ttg agc gag aaa tta cag gct tac ctg ggc cat gcg Ala Ala Glu Leu Ser Glu Lys Leu Gln Ala Tyr Leu Gly His Ala 3110 3115 3120	9369
att cgg gcc gac atc tat ctg agc gag gac acg ccc ggc aaa ccg Ile Arg Ala Asp Ile Tyr Leu Ser Glu Asp Thr Pro Gly Lys Pro 3125 3130 3135	9414
gca ggc gct ccg atc gtg gcc gag gaa gat ctg ctc acg ctg atg Ala Gly Ala Pro Ile Val Ala Glu Glu Asp Leu Leu Thr Leu Met 3140 3145 3150	9459
gat gcc tgg atc gaa aag ggc cag tac ggt cgt ttg ctg gag tac Asp Ala Trp Ile Glu Lys Gly Gln Tyr Gly Arg Leu Leu Glu Tyr 3155 3160 3165	9504
tgg acc aag ggc caa ccg atc gac tgg aac aaa ctc tat tgg cgc Trp Thr Lys Gly Gln Pro Ile Asp Trp Asn Lys Leu Tyr Trp Arg 3170 3175 3180	9549
aag ctg tat gcg gac gga cgg ccg cgg cgg atc agc ctg ccc acc Lys Leu Tyr Ala Asp Gly Arg Pro Arg Arg Ile Ser Leu Pro Thr 3185 3190 3195	9594

3185	3190	3195	
tat ccg ttc gag cac ccg cgt	tat tgg caa acg ccg	gtg ccg ggc	9639
Tyr Pro Phe Glu His Arg Arg	Tyr Trp Gln Thr Pro	Val Pro Gly	
3200	3205	3210	
gag cga agc ctg cac gcc acc	gcg cca gct act ccg	gaa acg gtt	9684
Glu Arg Ser Leu His Ala Thr	Ala Pro Ala Thr Arg	Glu Thr Val	
3215	3220	3225	
gcg gtt ggt gcc atg ccg gat	ccg gcc ggc gct acg	gtg caa gcc	9729
Ala Val Gly Ala Met Pro Asp	Pro Ala Gly Ala Thr	Val Gln Ala	
3230	3235	3240	
ccg ttg tgc gcc ttg tgc caa	gtg ttg ttg ggc aaa	ccg gtc acg	9774
Arg Leu Cys Ala Leu Cys Gln	Val Leu Leu Gly Lys	Pro Val Thr	
3245	3250	3255	
gcc cag atg gat ttc ttt gcc	gtc ggc ggc cat tcg	gtg ctg gcg	9819
Ala Gln Met Asp Phe Phe Ala	Val Gly Gly His Ser	Val Leu Ala	
3260	3265	3270	
atc caa ttg gtc tcg cgc atc	cgc aaa agc ttc ggg	gtg gag tat	9864
Ile Gln Leu Val Ser Arg Ile	Arg Lys Ser Phe Gly	Val Glu Tyr	
3275	3280	3285	
ccg gtc agc gct ttg ttc gaa	tcg gcg ctg ttg tcg	gac atg gcg	9909
Pro Val Ser Ala Leu Phe Glu	Ser Ala Leu Leu Ser	Asp Met Ala	
3290	3295	3300	
ccg cag atc gaa caa ttg ccg	gtg aac gga gtc gcc	aag cgc atg	9954
Arg Gln Ile Glu Gln Leu Arg	Val Asn Gly Val Ala	Lys Arg Met	
3305	3310	3315	
ccg gcg ttg ttg cct gcc ggg	cgc gtg ggc gcg att	cct gcg act	9999
Pro Ala Leu Leu Pro Ala Gly	Arg Val Gly Ala Ile	Pro Ala Thr	
3320	3325	3330	
tat gca cag gag cgc cta tgg	ctc gtc cac gaa cat	atg agt gag	10044
Tyr Ala Gln Glu Arg Leu Trp	Leu Val His Glu His	Met Ser Glu	
3335	3340	3345	
caa cgc agt agt tac aac atc	acc ttt gcc atg cac	ttc aga ggc	10089
Gln Arg Ser Ser Tyr Asn Ile	Thr Phe Ala Met His	Phe Arg Gly	
3350	3355	3360	
gtc gac ttc cgt gct gaa gcg	atg cgt gcc gca ttg	aac gcg ctg	10134
Val Asp Phe Arg Ala Glu Ala	Met Arg Ala Ala Leu	Asn Ala Leu	
3365	3370	3375	
gtg gtg ccg cac gaa gtg ctg	cgc aca cgc ttt ctt	tcg gag gac	10179
Val Val Arg His Glu Val Leu	Arg Thr Arg Phe Leu	Ser Glu Asp	
3380	3385	3390	
ggg cag ctg caa cag gtg atc	gct gcc tcg ttg acg	ttg gag gtg	10224
Gly Gln Leu Gln Gln Val Ile	Ala Ala Ser Leu Thr	Leu Glu Val	
3395	3400	3405	
ccg gta aga gag atg tcg gtc	gag gag gtc gac ctg	ctg ctg gcc	10269
Pro Val Arg Glu Met Ser Val	Glu Glu Val Asp Leu	Leu Leu Ala	
3410	3415	3420	
gcg agc acg ccg gag act ttc	gat ctg ccg cag ggg	ccc ttg ttc	10314

Ala Ser Thr Arg Glu Thr Phe Asp Leu Arg Gln Gly Pro Leu Phe	
3425 3430 3435	
aag gca cgc atc ctg cgc gtg gcg gcc gat cac cat gtg gtg ttg	10359
Lys Ala Arg Ile Leu Arg Val Ala Ala Asp His His Val Val Leu	
3440 3445 3450	
agc agc atc cac cac atc att tcc gac ggc tgg tcg ctg gga gtg	10404
Ser Ser Ile His His Ile Ile Ser Asp Gly Trp Ser Leu Gly Val	
3455 3460 3465	
ttc aac cgt gac ctg cac cag ctg tac gag gcg tgt ttg cgc ggc	10449
Phe Asn Arg Asp Leu His Gln Leu Tyr Glu Ala Cys Leu Arg Gly	
3470 3475 3480	
acg ccc ccc aca ctg ccg acg ctg gcg gtg cag tat gcc gac tac	10494
Thr Pro Pro Thr Leu Pro Thr Leu Ala Val Gln Tyr Ala Asp Tyr	
3485 3490 3495	
gcg ctg tgg caa cgg caa tgg gag ctg gcg gct ccg ctg tcg tac	10539
Ala Leu Trp Gln Arg Gln Trp Glu Leu Ala Ala Pro Leu Ser Tyr	
3500 3505 3510	
tgg acg cgg gca ctg gaa ggc tac gac gac ggc ctg gac ttg ccc	10584
Trp Thr Arg Ala Leu Glu Gly Tyr Asp Asp Gly Leu Asp Leu Pro	
3515 3520 3525	
tac gac cgg ccg cgc ggc gcc acg cgg gcg tgg cgg gca ggg ctg	10629
Tyr Asp Arg Pro Arg Gly Ala Thr Arg Ala Trp Arg Ala Gly Leu	
3530 3535 3540	
gtc aaa cac cgc tat ccg ccg caa ctg gcc cag cag ttg gcg gcc	10674
Val Lys His Arg Tyr Pro Pro Gln Leu Ala Gln Gln Leu Ala Ala	
3545 3550 3555	
tac agc caa cag tac caa gcg acg ctg ttc atg agc ctg ctg gca	10719
Tyr Ser Gln Gln Tyr Gln Ala Thr Leu Phe Met Ser Leu Leu Ala	
3560 3565 3570	
ggc ctg gcg ttg gtg ctg ggc cgt tac gcc gat cgc aag gac gtg	10764
Gly Leu Ala Leu Val Leu Gly Arg Tyr Ala Asp Arg Lys Asp Val	
3575 3580 3585	
tgc atc ggc gcg acg gtc tcc ggc cgc gac cag ctg gag ctg gaa	10809
Cys Ile Gly Ala Thr Val Ser Gly Arg Asp Gln Leu Glu Leu Glu	
3590 3595 3600	
gag ctg atc ggc ttt ttc atc aat att ttg ccg ctg cgg gtg gac	10854
Glu Leu Ile Gly Phe Phe Ile Asn Ile Leu Pro Leu Arg Val Asp	
3605 3610 3615	
ctg tcg ggg gat ccg tgc ctg gag gag gtg ctg ctg cgc acg cgt	10899
Leu Ser Gly Asp Pro Cys Leu Glu Glu Val Leu Leu Arg Thr Arg	
3620 3625 3630	
caa gtg gta ctg gat ggc ttc gcg cac cag tcg gtg ccg ttc gag	10944
Gln Val Val Leu Asp Gly Phe Ala His Gln Ser Val Pro Phe Glu	
3635 3640 3645	
cac gtg ttg cag gcg ctg cgg cgt cag cgc gac agt agc cag atc	10989
His Val Leu Gln Ala Leu Arg Arg Gln Arg Asp Ser Ser Gln Ile	
3650 3655 3660	
ccg ctg gtg ccg gtg atg ctg cga cac cag aac ttc ccg acg cag	11034

- 62 -

gcc ccg acg ccg ctg ggg ctg agg gaa cat ccg ggc gac ctg gcg Ala Pro Thr Pro Leu Gly Leu Arg Glu His Pro Gly Asp Leu Ala 3905 3910 3915	11754
tgc gtg atg gtg acc tcc ggc tgc acc ggc cgg ccc aag ggc gtg Cys Val Met Val Thr Ser Gly Ser Thr Gly Arg Pro Lys Gly Val 3920 3925 3930	11799
atg gtg ccg tat gcg cag ctg cac aac tgg ctg cat gca ggc tgg Met Val Pro Tyr Ala Gln Leu His Asn Trp Leu His Ala Gly Trp 3935 3940 3945	11844
cag cgt tct gcg ttc gag gcc ggg gag cgg gtg ctg cag aag acc Gln Arg Ser Ala Phe Glu Ala Gly Glu Arg Val Leu Gln Lys Thr 3950 3955 3960	11889
tgc atc gcc ttt gcg gtg tgc gta aag gag ttg cta agc ggc ctg Ser Ile Ala Phe Ala Val Ser Val Lys Glu Leu Leu Ser Gly Leu 3965 3970 3975	11934
ctg gcg ggg gtg gaa cag gtg atg ctg ccg gac gag cag gtg aag Leu Ala Gly Val Glu Gln Val Met Leu Pro Asp Glu Gln Val Lys 3980 3985 3990	11979
gac agc ctg gcg ttg gcg cgg gcg att gag caa tgg cag gtg acg Asp Ser Leu Ala Leu Ala Arg Ala Ile Glu Gln Trp Gln Val Thr 3995 4000 4005	12024
cgg ctg tac cta gtg cca tgc cac ctg cag gcg ctg ctg gac gcg Arg Leu Tyr Leu Val Pro Ser His Leu Gln Ala Leu Leu Asp Ala 4010 4015 4020	12069
acg caa gga cga gac ggg cta ctg cac tgc ctg cgt cac gtg gtg Thr Gln Gly Arg Asp Gly Leu Leu His Ser Leu Arg His Val Val 4025 4030 4035	12114
acg gcg ggg gaa gcg ttg ccg tct gcg gtg cgc gaa acg gtg cgg Thr Ala Gly Glu Ala Leu Pro Ser Ala Val Arg Glu Thr Val Arg 4040 4045 4050	12159
gtg cgt ctg cca cag gtg cag cta tgg aac aac tat ggc tgc acg Val Arg Leu Pro Gln Val Gln Leu Trp Asn Asn Tyr Gly Cys Thr 4055 4060 4065	12204
gaa ctg aac gac gcg acc tac cat cgg tgc gat acg gtg gcg cca Glu Leu Asn Asp Ala Thr Tyr His Arg Ser Asp Thr Val Ala Pro 4070 4075 4080	12249
gga acg ttt gtg ccg atc ggc gca ccg atc gcc aac acc gag gta Gly Thr Phe Val Pro Ile Gly Ala Pro Ile Ala Asn Thr Glu Val 4085 4090 4095	12294
tac gtg ctg gac cgg cag ctg cgg cag gtg ccg atc ggg gtg atg Tyr Val Leu Asp Arg Gln Leu Arg Gln Val Pro Ile Gly Val Met 4100 4105 4110	12339
ggc gag ctg cac gta cac agc gtg ggg atg gcg cgc ggc tac tgg Gly Glu Leu His Val His Ser Val Gly Met Ala Arg Gly Tyr Trp 4115 4120 4125	12384
aac cgg ccg ggg ctg acg gcc tgc cgc ttc atc gcg cac ccg tat Asn Arg Pro Gly Leu Thr Ala Ser Arg Phe Ile Ala His Pro Tyr 4130 4135 4140	12429

agc gag	gag ccg ggc aca cgg	ctg tac aag acc ggt	gac atg gta	12474
Ser Glu	Glu Pro Gly Thr Arg	Leu Tyr Lys Thr Gly	Asp Met Val	
4145	4150	4155		
cgc cgg	ctg gcg gac ggg acg	ctg gaa tac ctg ggc	cga cag gac	12519
Arg Arg	Leu Ala Asp Gly Thr	Leu Glu Tyr Leu Gly	Arg Gln Asp	
4160	4165	4170		
ttc gag	gtc aag gtg cgc ggc	cac cgg gtg gat acg	cgg cag gtg	12564
Phe Glu	Val Lys Val Arg Gly	His Arg Val Asp Thr	Arg Gln Val	
4175	4180	4185		
gag gcg	gcc ttg cgg gcg cag	ccc gcg gtg gcc gag	gcg gtg gtg	12609
Glu Ala	Ala Leu Arg Ala Gln	Pro Ala Val Ala Glu	Ala Val Val	
4190	4195	4200		
agc ggt	cac cgg gtg gac ggg	gac atg cag ttg gtg	gcc tat gtg	12654
Ser Gly	His Arg Val Asp Gly	Asp Met Gln Leu Val	Ala Tyr Val	
4205	4210	4215		
gtg gcg	cgt gaa ggg cag gca	ccg agc gcg ggc gag	ttg aaa caa	12699
Val Ala	Arg Glu Gly Gln Ala	Pro Ser Ala Gly Glu	Leu Lys Gln	
4220	4225	4230		
cag ctg	tcg gcg cag ttg ccg	acc tac atg ctg ccg	acc gtg tac	12744
Gln Leu	Ser Ala Gln Leu Pro	Thr Tyr Met Leu Pro	Thr Val Tyr	
4235	4240	4245		
cag tgg	ctg gag cag ttg ccg	cgg ctg tcc aac ggc	aag ttg gac	12789
Gln Trp	Leu Glu Gln Leu Pro	Arg Leu Ser Asn Gly	Lys Leu Asp	
4250	4255	4260		
cgg ttg	gcg ctg ccg gcg ccg	cag gtg gta cac gcg	cag gag tac	12834
Arg Leu	Ala Leu Pro Ala Pro	Gln Val Val His Ala	Gln Glu Tyr	
4265	4270	4275		
gtc gcg	cca cgc aac gag gcc	gag caa cgg ctg gcg	gca ctg ttt	12879
Val Ala	Pro Arg Asn Glu Ala	Glu Gln Arg Leu Ala	Ala Leu Phe	
4280	4285	4290		
gcc gag	gtg ctg cgg gtg gag	cag gtg ggc atc cac	gac aac ttc	12924
Ala Glu	Val Leu Arg Val Glu	Gln Val Gly Ile His	Asp Asn Phe	
4295	4300	4305		
ttc gcc	ttg ggt ggg cac tcg	ctg tct gca tcg caa	ctg atc tcg	12969
Phe Ala	Leu Gly Gly His Ser	Leu Ser Ala Ser Gln	Leu Ile Ser	
4310	4315	4320		
cgc atc	cgc caa agt ttt cac	gtc gat ctg ccg ctg	agc cgg atc	13014
Arg Ile	Arg Gln Ser Phe His	Val Asp Leu Pro Leu	Ser Arg Ile	
4325	4330	4335		
ttc gag	gca ccc acg atc gag	ggc ctg gtc agg cag	cta gcg ttg	13059
Phe Glu	Ala Pro Thr Ile Glu	Gly Leu Val Arg Gln	Leu Ala Leu	
4340	4345	4350		
cct agt	gaa ggc ggc gtg gcc	agc atc gcc agg gta	gcg cga aac	13104
Pro Ser	Glu Gly Gly Val Ala	Ser Ile Ala Arg Val	Ala Arg Asn	
4355	4360	4365		
cgg acg	atc cca ttg tcg ctg	ttc cag gaa cgc ctg	tgg ttc gtg	13149
Arg Thr	Ile Pro Leu Ser Leu	Phe Gln Glu Arg Leu	Trp Phe Val	
4370	4375	4380		

cac caa	cac atg	cct gag	caa cgc	acc agt	tac aac	ggc acg	ctc	13194
His Gln	His Met	Pro Glu	Gln Arg	Thr Ser	Tyr Asn	Gly Thr	Leu	
4385			4390			4395		
gcc ttg	cgt ttg	cgt ggt	cct ttg	tcg gtg	gaa gcg	atg cgt	gca	13239
Ala Leu	Arg Leu	Arg Gly	Pro Leu	Ser Val	Glu Ala	Met Arg	Ala	
4400			4405			4410		
gcg ctg	cgt gcg	tta gtg	ctg cgc	cac gaa	atc ttg	cgt acc	cgc	13284
Ala Leu	Arg Ala	Leu Val	Leu Arg	His Glu	Ile Leu	Arg Thr	Arg	
4415			4420			4425		
ttc gtg	ttg ccg	acc ggt	gct agc	gag ccg	gtg cag	gtc att	gac	13329
Phe Val	Leu Pro	Thr Gly	Ala Ser	Glu Pro	Val Gln	Val Ile	Asp	
4430			4435			4440		
gag cac	agc gat	ttc cag	ctc tca	gta cag	cta gtc	gag gat	act	13374
Glu His	Ser Asp	Phe Gln	Leu Ser	Val Gln	Leu Val	Glu Asp	Thr	
4445			4450			4455		
gag atc	gcg tcg	ctg atg	gat gaa	ctg gca	agt cat	atc tac	gac	13419
Glu Ile	Ala Ser	Leu Met	Asp Glu	Leu Ala	Ser His	Ile Tyr	Asp	
4460			4465			4470		
tta gcc	aac ggc	ccg ctg	ttc att	gca tgc	ctt ttg	caa ctg	gat	13464
Leu Ala	Asn Gly	Pro Leu	Phe Ile	Ala Cys	Leu Leu	Gln Leu	Asp	
4475			4480			4485		
gag caa	gaa cat	gtg ctg	cta atc	ggc atg	cat cac	ctt atc	tac	13509
Glu Gln	Glu His	Val Leu	Leu Ile	Gly Met	His His	Leu Ile	Tyr	
4490			4495			4500		
gac gct	tgg tcg	caa ttc	acc gtg	atg aac	cgc gat	cta cgc	gtg	13554
Asp Ala	Trp Ser	Gln Phe	Thr Val	Met Asn	Arg Asp	Leu Arg	Val	
4505			4510			4515		
ctg tat	cac cgc	cac ctg	gga ctt	gcc ggc	gga gat	ctg ccg	gaa	13599
Leu Tyr	His Arg	His Leu	Gly Leu	Ala Gly	Gly Asp	Leu Pro	Glu	
4520			4525			4530		
tta ccg	atc caa	tat gcc	gac tat	gcg atc	tgg caa	cgc gcc	cag	13644
Leu Pro	Ile Gln	Tyr Ala	Asp Tyr	Ala Ile	Trp Gln	Arg Ala	Gln	
4535			4540			4545		
aac ctg	gac gcg	caa ctg	gcc tat	tgg cag	gct atg	ttg cac	gac	13689
Asn Leu	Asp Ala	Gln Leu	Ala Tyr	Trp Gln	Ala Met	Leu His	Asp	
4550			4555			4560		
tac gac	gac ggc	ctg gag	ctg ccc	tac gac	tat ccg	cgt ccg	cgc	13734
Tyr Asp	Asp Gly	Leu Glu	Leu Pro	Tyr Asp	Tyr Pro	Arg Pro	Arg	
4565			4570			4575		
aat cgc	acc tgg	cac gca	gcg gtc	tac aca	cac acc	tat ccg	gct	13779
Asn Arg	Thr Trp	His Ala	Ala Val	Tyr Thr	His Thr	Tyr Pro	Ala	
4580			4585			4590		
gaa ctg	gta cag	cgc ttt	gcc ggc	ttc gta	cag gcg	cat cag	tcg	13824
Glu Leu	Val Gln	Arg Phe	Ala Gly	Phe Val	Gln Ala	His Gln	Ser	
4595			4600			4605		
acc ttg	ttc atc	ggg ctg	ttg gcc	agc ttc	gcg gtc	gtg ttg	aac	13869
Thr Leu	Phe Ile	Gly Leu	Leu Ala	Ser Phe	Ala Val	Val Leu	Asn	

4610	4615	4620	
aaa tac acc ggc cgg gac gac	ttg tgc atc ggt acc	acc acg gca	13914
Lys Tyr Thr Gly Arg Asp Asp	Leu Cys Ile Gly Thr	Thr Thr Ala	
4625	4630	4635	
ggg cgc acc cac ctg gag ctg	gag aac ctg atc ggt	ttc ttc atc	13959
Gly Arg Thr His Leu Glu Leu	Glu Asn Leu Ile Gly	Phe Phe Ile	
4640	4645	4650	
aac atc ttg cct ttg cgc ttg	cgc ttg gac ggc gat	ccg gac gtt	14004
Asn Ile Leu Pro Leu Arg Leu	Arg Leu Asp Gly Asp	Pro Asp Val	
4655	4660	4665	
gcc gaa atc atg cgg cga aca	cgg ttg gtg gcg atg	agc gcg ttt	14049
Ala Glu Ile Met Arg Arg Thr	Arg Leu Val Ala Met	Ser Ala Phe	
4670	4675	4680	
gag aac cag gcg cta ccg ttc	gag cac ctg ctc aac	gcc ctg cac	14094
Glu Asn Gln Ala Leu Pro Phe	Glu His Leu Leu Asn	Ala Leu His	
4685	4690	4695	
aag caa cgt gac acc agc cgg	att ccg cta gtt ccg	gtg gtg atg	14139
Lys Gln Arg Asp Thr Ser Arg	Ile Pro Leu Val Pro	Val Val Met	
4700	4705	4710	
cgt cat cag aac ttc ccg gac	acg atc ggc gac tgg	agc gat ggc	14184
Arg His Gln Asn Phe Pro Asp	Thr Ile Gly Asp Trp	Ser Asp Gly	
4715	4720	4725	
atc cgt acc gaa gtg atc cag	cgc gat ctg cgt gcc	acc ccc aat	14229
Ile Arg Thr Glu Val Ile Gln	Arg Asp Leu Arg Ala	Thr Pro Asn	
4730	4735	4740	
gaa atg gac ctg caa ttc ttc	ggc gac ggt acg ggg	ctt tcg gtc	14274
Glu Met Asp Leu Gln Phe Phe	Gly Asp Gly Thr Gly	Leu Ser Val	
4745	4750	4755	
aca gtg gaa tac gcg gcg gag	ctg ttc tca gaa gcg	acc att cgc	14319
Thr Val Glu Tyr Ala Ala Glu	Leu Phe Ser Glu Ala	Thr Ile Arg	
4760	4765	4770	
cgc ctg atc cac cat cac caa	ctc gtc ctg gag cag	atg ttg gcg	14364
Arg Leu Ile His His His Gln	Leu Val Leu Glu Gln	Met Leu Ala	
4775	4780	4785	
gcc cat gaa agc gcc acg tgc	ccc ttg gat gtt gcc	gac tag	14406
Ala His Glu Ser Ala Thr Cys	Pro Leu Asp Val Ala	Asp	
4790	4795	4800	

<210> 4
 <211> 4801
 <212> PRT
 <213> Xanthomonas albilineans

<220>
 <221> misc_feature
 <222> (226)..(240)
 <223> Acyl-CoA ligase subdomain I

<220>
 <221> misc_feature

<222> (486)..(493)
<223> Acyl-CoA ligase subdomain II

<220>
<221> misc_feature
<222> (897)..(913)
<223> Beta-ketoacyl synthase 1 subdomain I

<220>
<221> misc_feature
<222> (1038)..(1047)
<223> Beta-ketoacyl synthase 1 subdomain II

<220>
<221> misc_feature
<222> (1080)..(1089)
<223> Beta-ketoacyl synthase 1 subdomain III

<220>
<221> misc_feature
<222> (2777)..(2793)
<223> Beta-ketoacyl synthase 2 subdomain I

<220>
<221> misc_feature
<222> (2918)..(2927)
<223> Beta-ketoacyl synthase 2 subdomain II

<220>
<221> misc_feature
<222> (2955)..(2964)
<223> Beta-ketoacyl synthase 2 subdomain III

<220>
<221> misc_feature
<222> (1912)..(1942)
<223> Beta-ketoacyl reductase domain

<220>
<221> misc_feature
<222> (667)..(678)
<223> Acyl carrier protein 1 domain

<220>
<221> misc_feature
<222> (2484)..(2495)
<223> Acyl carrier protein 2 domain

<220>
<221> misc_feature
<222> (2568)..(2579)
<223> Acyl carrier protein 3 domain

<220>
<221> misc_feature
<222> (3806)..(3811)
<223> Adenylation subdomain I

<220>
<221> misc_feature
<222> (3851)..(3861)
<223> Adenylation subdomain II

<220>
<221> misc_feature
<222> (3917)..(3932)
<223> Adenylation subdomain III

<220>
<221> misc_feature
<222> (3967)..(3970)
<223> Adenylation subdomain IV

<220>
<221> misc_feature
<222> (4063)..(4069)
<223> Adenylation subdomain V

<220>
<221> misc_feature
<222> (4114)..(4128)
<223> Adenylation subdomain VI

<220>
<221> misc_feature
<222> (4152)..(4157)
<223> Adenylation subdomain VII

<220>
<221> misc_feature
<222> (4170)..(4189)
<223> Adenylation subdomain VIII

<220>
<221> misc_feature
<222> (4239)..(4245)
<223> Adenylation subdomain IX

<220>
<221> misc_feature
<222> (4259)..(4264)
<223> Adenylation subdomain X

<220>
<221> misc_feature
<222> (3261)..(3271)
<223> Peptidyl carrier protein 1 domain

<220>
<221> misc_feature
<222> (4306)..(4316)
<223> Peptidyl carrier protein 2 domain

<220>
<221> misc_feature
<222> (3333)..(3342)
<223> Condensation domain 1 subdomain I

<220>
<221> misc_feature
<222> (3381)..(3389)
<223> Condensation domain 1 subdomain II

<220>
<221> misc_feature
<222> (3456)..(3465)

```

<223> Condensation domain 1 subdomain III

<220>
<221> misc_feature
<222> (3495)..(3501)
<223> Condensation domain 1 subdomain IV

<220>
<221> misc_feature
<222> (3606)..(3617)
<223> Condensation domain 1 subdomain V

<220>
<221> misc_feature
<222> (3641)..(3647)
<223> Condensation domain 1 subdomain VI

<220>
<221> misc_feature
<222> (3658)..(3665)
<223> Condensation domain 1 subdomain VII

<220>
<221> misc_feature
<222> (4374)..(4383)
<223> Condensation domain 2 subdomain I

<220>
<221> misc_feature
<222> (4421)..(4429)
<223> Condensation domain 2 subdomain II

<220>
<221> misc_feature
<222> (4498)..(4507)
<223> Condensation domain 2 subdomain III

<220>
<221> misc_feature
<222> (4538)..(4544)
<223> Condensation domain 2 subdomain IV

<220>
<221> misc_feature
<222> (4649)..(4659)
<223> Condensation domain 2 subdomain V

<220>
<221> misc_feature
<222> (4685)..(4691)
<223> Condensation domain 2 subdomain VI

<220>
<221> misc_feature
<222> (4701)..(4708)
<223> Condensation domain 2 subdomain VII

<400> 4

```

```

Leu Pro Asn Ala Leu Met Gln Ile Thr Leu Val Ala Val Gln Phe Ala
1           5           10          15

```

```

Gly Val Leu Leu Gly Val Thr Ala Arg Ala Ala Ile Pro Asn Lys Ala

```

20 25 30
 Gly Met Arg Arg Ala Trp Pro Pro Phe Pro Gln Ala Cys Cys Arg Ser
 35 40 45
 Ile Ala Tyr Leu Met Gln Arg Ser Pro Met Ser Pro Leu Gln Gln Thr
 50 55 60
 Leu Leu Thr Arg Leu Ala Ser Ala Ala Ser Arg Thr Met Ile Glu
 65 70 75 80
 Phe Pro Arg Pro Glu His Ala Ser Pro Gln Cys Cys Asp Asp Ala Glu
 85 90 95
 Leu Ala Arg Leu Ile Val Gln Leu Ser Ala Gly Leu Gln Pro Leu Ala
 100 105 110
 Met Pro Gly Thr Tyr Val Ile Ile Ala Ala Pro His Gly Gly Leu Phe
 115 120 125
 Ala Ala Ala Leu Leu Ala Cys Leu His Ala Asn Leu Val Ala Val Pro
 130 135 140
 Phe Pro Leu Asp Val Ala Gln Pro Asn Glu Arg Glu Gln Ala Arg Leu
 145 150 155 160
 Glu Thr Ile His Ala Gln Leu Met Glu His Gly Asn Val Ala Val Leu
 165 170 175
 Leu Asp Asp Val Ala Asp Arg Ser Ala Phe Ala Arg Met Ala His Ala
 180 185 190
 Ala Gly Thr Phe Leu Ala Thr Phe Ala Asp Leu Lys Arg Glu Ser Thr
 195 200 205
 Ser Ala Ser Leu Cys Pro Ala Ser Pro Ser Asp Ala Ala Leu Leu Leu
 210 215 220
 Phe Thr Ser Gly Ser Ser Gly Glu Ser Lys Gly Ile Leu Leu Ser His
 225 230 235 240
 Arg Asn Leu His His Gln Ile Gln Ala Gly Ile Arg Gln Trp Ser Leu
 245 250 255
 Asp Glu His Ser His Val Val Thr Trp Leu Ser Pro Ala His Asn Phe
 260 265 270
 Gly Leu His Phe Gly Leu Leu Ala Pro Trp Phe Ser Gly Ala Thr Val

275

280

285

Ser Phe Ile His Pro His Ser Tyr Met Lys Arg Pro Gly Phe Trp Leu
 290 295 300

Glu Thr Val Ala Ala Arg Asp Ala Thr His Met Ala Ala Pro Asn Phe
 305 310 315 320

Ala Phe Asp Tyr Cys Cys Asp Trp Val Met Val Glu Gln Leu Pro Pro
 325 330 335

Ser Ala Leu Ser Thr Leu Thr His Ile Val Cys Gly Gly Glu Pro Val
 340 345 350

Arg Ala Ser Thr Met Gln Arg Phe Phe Glu Lys Phe Ala Gly Leu Gly
 355 360 365

Ala Arg Thr Gln Thr Phe Met Pro His Phe Gly Leu Ser Glu Thr Gly
 370 375 380

Ala Leu Ser Thr Leu Asp Glu Ala Pro Gln Gln Arg Val Leu Glu Leu
 385 390 395 400

Asp Ala Asp Ala Leu Asn Lys Arg Lys Arg Val Ala Ala Gly Ala Ser
 405 410 415

Glu Ala Arg Val Thr Val Leu Asn Lys Gly Ala Val Asp Glu Asp Val
 420 425 430

Glu Leu Arg Ile Val Cys Pro Glu Gly Glu Thr Leu Cys Arg Pro Asp
 435 440 445

Glu Ile Gly Glu Ile Trp Val Lys Ser Pro Ala Ile Ala Arg Gly Tyr
 450 455 460

Leu Phe Ala Lys Pro Ala Asp Gln Arg Gln Phe Asn Cys Ser Ile Arg
 465 470 475 480

His Thr Asp Asp Ser Gly Tyr Phe Arg Thr Gly Asp Leu Gly Phe Ile
 485 490 495

Ala Asp Gly Cys Leu Tyr Val Thr Gly Arg Val Lys Glu Val Leu Ile
 500 505 510

Ile Arg Gly Lys Asn His Tyr Pro Ala His Ile Glu Ala Ser Ile Ala
 515 520 525

Ala Thr Ala Ser Pro Gly Ala Leu Met Pro Val Val Phe Ser Ile Glu
530 535 540

Arg Gln Asp Glu Glu Arg Val Ala Ala Val Ile Ala Val Asn His Pro
545 550 555 560

Trp Thr Pro Ala Ala Cys Ala Ala Gln Ala His Lys Ile Arg Gln Gln
565 570 575

Val Ala Asp Gln His Gly Val Ala Leu Ala Glu Leu Ala Phe Ala Glu
580 585 590

His Arg His Val Phe Gly Thr Tyr Pro Gly Lys Leu Lys Arg Arg Leu
595 600 605

Val Lys Glu Ala Tyr Val Asn Gly Gln Leu Pro Leu Leu Trp His Glu
610 615 620

Gly Lys Asn Arg Asp Val Pro Ala Ala Ala Ala Asp Asp Arg Gln Ala
625 630 635 640

Gln His Val Ala Asp Leu Cys Arg Lys Val Phe Leu Pro Val Leu Gly
645 650 655

Val Ala Pro Pro His Ala Gln Trp Pro Leu Cys Glu Leu Ala Leu Asp
660 665 670

Ser Leu Gln Cys Val Arg Leu Ala Gly Ala Ile Glu Glu Cys Tyr Gly
675 680 685

Val Pro Phe Glu Pro Thr Leu Leu Phe Lys Leu Glu Thr Val Gly Ala
690 695 700

Ile Ala Glu Tyr Val Leu Ala His Gly Arg Gln Ala Pro Thr Pro Thr
705 710 715 720

Arg Ala Pro Val Ala Ser Thr Thr Cys Ser Glu Glu Pro Ile Ala Ile
725 730 735

Val Ala Met His Cys Glu Val Pro Gly Ala Gly Glu Asn Thr Glu Ala
740 745 750

Leu Trp Ser Phe Leu Arg Ser Asp Val Asn Ala Ile Arg Pro Ile Glu
755 760 765

Ser Thr Arg Pro Asp Leu Trp Ala Ala Met Arg Ala Tyr Pro Gly Leu
770 775 780

Ala Gly Glu Gln Leu Pro Arg Tyr Ala Gly Phe Leu Asp Asp Val Asp
785 790 795 800

Ala Phe Asp Ala Ala Phe Phe Gly Ile Ser Arg Arg Glu Ala Glu Cys
805 810 815

Met Asp Pro Gln Gln Arg Lys Val Leu Glu Met Val Trp Lys Leu Ile
820 825 830

Glu Gln Ala Gly His Asp Pro Leu Ser Trp Gly Gly Gln Pro Val Gly
835 840 845

Leu Phe Val Gly Ala His Thr Ser Asp Tyr Gly Glu Leu Leu Ala Ser
850 855 860

Gln Pro Gln Leu Met Ala Gln Cys Gly Ala Tyr Ile Asp Ser Gly Ser
865 870 875 880

His Leu Thr Met Ile Pro Asn Arg Ala Ser Arg Trp Phe Asn Phe Thr
885 890 895

Gly Pro Ser Glu Val Ile Asn Ser Ala Cys Ser Ser Ser Leu Val Ala
900 905 910

Leu His Arg Ala Val Gln Ser Leu Arg Gln Gly Glu Ser Ser Val Ala
915 920 925

Leu Val Leu Gly Val Asn Leu Ile Leu Ala Pro Lys Val Leu Leu Ala
930 935 940

Ser Ala Ser Ala Gly Met Leu Ser Pro Asp Gly Arg Cys Lys Thr Leu
945 950 955 960

Asp Ala Ala Ala Asp Gly Phe Val Arg Ser Glu Gly Ile Ala Gly Val
965 970 975

Ile Leu Lys Pro Leu Ala Gln Ala Leu Ala Asp Gly Asp Arg Val Tyr
980 985 990

Gly Leu Val Arg Gly Val Ala Val Asn His Gly Gly Arg Ser Asn Ser
995 1000 1005

Leu Arg Ala Pro Asn Val Asn Ala Gln Arg Gln Leu Leu Ile Arg
1010 1015 1020

Thr Tyr Gln Glu Ala Gly Val Glu Pro Ala Ser Val Gly Tyr Val
1025 1030 1035

Glu Leu His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu Ile
 1040 1045 1050

Gln Ala Leu Lys Glu Ala Phe Ile Ala Leu Gly Ala Gln Ala Ala
 1055 1060 1065

Pro Ser Asn Cys Gly Ile Gly Ser Val Lys Ser Ala Leu Gly His
 1070 1075 1080

Leu Glu Ala Ala Ala Gly Leu Thr Gly Leu Ile Lys Val Leu Leu
 1085 1090 1095

Met Leu Lys His Gly Glu Gln Ala Gly Thr Arg His Phe Ser Thr
 1100 1105 1110

Leu Asn Pro Leu Ile Asp Leu Arg Gly Thr Ser Phe Glu Val Val
 1115 1120 1125

Ala Gln His Arg Ala Trp Pro Ser Gln Val Gly Ile His Gly Thr
 1130 1135 1140

Leu Leu Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly Gly
 1145 1150 1155

Ala Asn Ala His Ala Ile Val Glu Glu His Val Ile Ala Thr Pro
 1160 1165 1170

Pro Ser Thr Ser Ser Ala Gly Gly Pro Val Gly Ile Val Leu Ser
 1175 1180 1185

Ala Gly Ser Glu Ala Val Leu Arg Gln Gln Val Leu Ala Leu Ser
 1190 1195 1200

Ala Trp Leu Arg Gln Gln Ser Pro Thr Pro Ala Gln Met Ile Asp
 1205 1210 1215

Val Ala Tyr Thr Leu Gln Val Gly Arg Ala Ala Leu Ser His Arg
 1220 1225 1230

Leu Ala Phe Ser Ala Thr Asp Ala Glu Gln Ala Leu Ala Arg Leu
 1235 1240 1245

Glu Gly Arg Leu Ala Gly Val Met Asp Ala Glu Val His His Gly
 1250 1255 1260

Val Val Asp Ala Ala Ala Thr Ala Pro Glu His Gly Arg Gln Thr
 1265 1270 1275

Arg Glu Gly Leu Ala Gly Leu Leu Arg Ala Trp Thr Gln Gly Val
 1280 1285 1290
 Arg Val Asp Trp Ser Ala Leu Tyr Gly Ile Gln Arg Pro Gln Arg
 1295 1300 1305
 Val Ser Leu Pro Val Tyr Pro Phe Ala Arg Glu Arg Tyr Trp Leu
 1310 1315 1320
 Pro Gly Gln Ala Met His Ala Ala Ala Asp Ala His Pro Met Leu
 1325 1330 1335
 Gln Leu Leu His Ala Asn Ala Lys Leu His Arg Tyr Ala Leu Arg
 1340 1345 1350
 Arg Ser Gly Cys Ala Ser Phe Leu Val Asp His Cys Val Asp Gly
 1355 1360 1365
 Arg Gln Val Leu Pro Ala Ala Val Gln Leu Glu Leu Val Arg Ala
 1370 1375 1380
 Val Ala Gln Arg Val Met Ala Gln Asp Glu Gly Cys Ile Glu Leu
 1385 1390 1395
 Ala Gln Val Ala Phe Leu His Pro Leu Met Met Glu Glu Thr Glu
 1400 1405 1410
 Leu Glu Val Glu Ile Glu Leu Ser Lys Ser Asp Gln Asp Glu Phe
 1415 1420 1425
 Asp Phe Gln Leu His Asp Ala His Arg Gln Gln Val Phe Ser Gln
 1430 1435 1440
 Gly His Val Arg Arg Arg Val Tyr Thr Ala Thr Pro Arg Leu Asp
 1445 1450 1455
 Leu Ala Gln Leu Gln Lys Leu Cys Ala Glu Arg Val Leu Ser Gly
 1460 1465 1470
 Glu Asp Cys Tyr Ala His Phe Thr Ala Cys Gly Leu Gln Leu Gly
 1475 1480 1485
 Asp Arg Leu Lys Ser Val Gln Ser Ile Gly Cys Gly Arg Asn Gly
 1490 1495 1500
 Glu Gly Glu Pro Ile Ala Leu Gly Val Leu Arg Leu Pro Pro Ser

1505	1510	1515
Ser Val Glu Asp Ser His Val	Leu Pro Pro Ser Leu	Leu Asp Gly
1520	1525	1530
Ala Leu Gln Cys Ser Leu Gly	Leu Gln Arg Asp Val	Glu His Ile
1535	1540	1545
Ala Met Pro Tyr Thr Leu Glu	Arg Met Thr Val His	Ala Pro Ile
1550	1555	1560
Pro Pro Glu Ala Trp Val Leu	Leu Arg His Gly His	Ala Ala Arg
1565	1570	1575
Gln Ser Leu Asp Ile Asp Leu	Leu Asp Ser Glu Gly	Arg Val Cys
1580	1585	1590
Val Ser Leu Gly Asn Tyr Thr	Gly Arg Ala Pro Lys	Ala Val Ser
1595	1600	1605
Ala Val Arg Ala Leu Val Leu	Ala Pro Val Trp Gln	Ala Leu Thr
1610	1615	1620
Glu Thr Ala Pro Ala Trp Pro	Asp Pro Ala Glu Arg	Ile Val Thr
1625	1630	1635
Val Gly Asp Asp Ala Trp Arg	Ser His Phe Gly Phe	Asp Glu Pro
1640	1645	1650
Ala Leu Ser Leu Glu Asp Ser	Val Glu Val Ile Ala	Thr Arg Leu
1655	1660	1665
Gly Gln Ser Gly Lys Phe Asp	His Leu Val Trp Ile	Val Pro Ile
1670	1675	1680
Ala Glu Ser Glu Thr Asp Ile	Ala Ala Gln Gly Ser	Ala Ala Ile
1685	1690	1695
Ala Gly Phe Arg Leu Val Lys	Ala Leu Leu Ala Leu	Gly Tyr Ala
1700	1705	1710
His Arg Pro Leu Gly Leu Thr	Val Leu Thr Arg Gln	Ala Leu Thr
1715	1720	1725
Arg Gln Pro Ser His Ala Ala	Val His Gly Leu Ile	Gly Thr Leu
1730	1735	1740
Ala Lys Glu Tyr Cys Asn Trp	Lys Ile Arg Leu Leu	Asp Leu Pro

1745	1750	1755
Ser Val Lys Ser Trp Pro Gln	Trp Glu Gln Leu Arg	Ser Leu Pro
1760	1765	1770
Trp His Ala Gln Gly Glu Ala	Leu Ile Gly Arg Gly	Thr Cys Trp
1775	1780	1785
Tyr Arg Arg Gln Leu Cys Glu	Val Leu Pro Leu Pro	Ser Leu Glu
1790	1795	1800
Pro Pro Pro Tyr Arg Val Gly	Gly Val Tyr Val Val	Ile Gly Gly
1805	1810	1815
Ala Gly Gly Leu Gly Glu Val	Leu Ser Glu His Leu	Ile Arg Thr
1820	1825	1830
Tyr Asp Ala Gln Leu Ile Trp	Ile Gly Arg Arg Val	Leu Asp Glu
1835	1840	1845
Gly Ile Ala Arg Lys Gln Thr	Arg Leu Ala Ser Leu	Gly Arg Ala
1850	1855	1860
Pro His Tyr Ile Ser Ala Asp	Ala Ser Asp Pro Ala	Ala Leu Gln
1865	1870	1875
Ala Ala His Asn Glu Ile Val	Ala Leu His Gly Gln	Pro His Gly
1880	1885	1890
Leu Ile Leu Ser Asn Ile Val	Leu Lys Asp Ala Ser	Leu Ala Arg
1895	1900	1905
Met Glu Glu Ala Asp Phe Arg	Asp Val Leu Ala Ala	Lys Leu Asp
1910	1915	1920
Val Ser Val Cys Ala Ala Gln	Val Phe Gly Thr Ala	Pro Leu Asp
1925	1930	1935
Phe Val Leu Phe Phe Ser Ser	Ile Gln Ser Thr Thr	Lys Ala Ala
1940	1945	1950
Gly Gln Gly Asn Tyr Ala Ala	Gly Cys Cys Tyr Val	Asp Ala Phe
1955	1960	1965
Gly Glu Leu Trp Ala Arg Arg	Gly Leu Arg Val Lys	Thr Ile Asn
1970	1975	1980

Trp Gly	Tyr Trp Gly Ser Val	Gly Val Val Ala Gly	Glu Asp Tyr
1985	1990	1995	
Arg Arg	Arg Met Ala Gln Lys	His Met Ala Ser Ile	Glu Gly Ala
2000	2005	2010	
Glu Ala	Met Gln Val Leu Ser	Gln Leu Leu Cys Ala	Pro Leu Gln
2015	2020	2025	
Arg Leu	Ala Tyr Val Lys Ile	Asp Asp Ala Asn Ala	Met Arg Ala
2030	2035	2040	
Leu Gly	Val Val Glu Asp Glu	Ser Val Gln Ile Pro	Val His Ala
2045	2050	2055	
Pro Ala	Glu Pro Pro Arg Gly	Gln Pro Gly Pro Val	Val Glu Leu
2060	2065	2070	
Ser Val	Asn Leu Asp Ala Arg	Arg Glu Arg Glu Thr	Leu Leu Ala
2075	2080	2085	
Ala Trp	Leu Leu Glu Leu Ile	Glu Gln Leu Gly Gly	Phe Pro Pro
2090	2095	2100	
Ala Ser	Phe Asp Ile Ala Thr	Leu Ala Gln Arg Leu	His Ile Val
2105	2110	2115	
Pro Ala	Tyr Arg Ser Trp Leu	Glu His Ser Val Arg	Met Leu Gly
2120	2125	2130	
Val Tyr	Gly Tyr Leu Arg Ala	Thr Gly Glu Ser Arg	Phe Glu Leu
2135	2140	2145	
Ala Asp	Lys Pro Pro Asp Asp	Ala Arg Gly Ala Trp	Asn Ala His
2150	2155	2160	
Val His	Glu Ala Ser Val Glu	Ala Gly Glu Glu Ala	Gln Arg Arg
2165	2170	2175	
Leu Leu	Asp Arg Cys Met Arg	Ala Leu Pro Ala Val	Leu Arg Gly
2180	2185	2190	
Glu Arg	Lys Ala Thr Glu Leu	Leu Phe Pro Glu Gly	Ser Met Ala
2195	2200	2205	
Trp Val	Glu Gly Ile Tyr Gln	Asn Asn Pro Leu Ala	Asp Tyr Phe
2210	2215	2220	

Asn Ala	Gln Leu Val Thr Arg	Leu Ile Ala Tyr Leu	Arg Arg Arg
2225	2230	2235	
Leu Glu	Ser Thr Pro Thr Ala	Arg Leu Lys Leu Cys	Glu Ile Gly
2240	2245	2250	
Ala Gly	Ser Gly Gly Thr Thr	Ala Ser Val Leu Gln	Gln Leu Gln
2255	2260	2265	
Ala Tyr	Gly Glu His Ile Glu	Glu Tyr Leu Tyr Thr	Asp Leu Ser
2270	2275	2280	
Pro Val	Phe Leu His His Ala	Glu Lys His Tyr Gln	Pro Arg Ala
2285	2290	2295	
Pro Tyr	Leu Arg Thr Ala Cys	Phe Asp Val Ala Arg	Ala Pro Thr
2300	2305	2310	
Ala Gln	Ala Leu Glu Ser Gly	Gly Tyr Asp Val Val	Ile Ala Ala
2315	2320	2325	
Asn Val	Leu His Ala Thr Arg	Asp Ile Ala Lys Thr	Leu Arg Asn
2330	2335	2340	
Ala Lys	Ala Leu Leu Lys Pro	Gly Gly Leu Leu Leu	Leu Asn Glu
2345	2350	2355	
Val Ile	Glu Arg Ser Leu Val	Leu His Leu Thr Phe	Gly Leu Leu
2360	2365	2370	
Glu Ser	Trp Trp Leu Pro Gln	Asp Lys Ile Leu Arg	Leu Ala Gly
2375	2380	2385	
Ser Pro	Leu Leu Ala Cys Ala	Thr Trp Arg Ser Leu	Leu Glu Ala
2390	2395	2400	
Glu Gly	Phe Ala Gly Leu Ser	Val His Arg Ala Gln	Pro Asp Ala
2405	2410	2415	
Gly Gln	Ala Ile Ile Cys Ala	Tyr Ser Asp Gly Ile	Val Arg Gln
2420	2425	2430	
Ala Ser	Thr Ile Glu Val Ala	Arg Asn Glu Lys Val	Thr Val Pro
2435	2440	2445	
Ser Gln	Pro Ala Glu Ala Gly	Glu Ser Pro Leu Asp	Leu Val Lys
2450	2455	2460	

Lys	Leu	Leu	Gly	Arg	Ile	Leu	Lys	Met	Asp	Pro	Ala	Thr	Leu	Asp
2465						2470					2475			
Thr	Ser	His	Pro	Leu	Glu	Tyr	Tyr	Gly	Val	Asp	Ser	Ile	Val	Ala
2480						2485					2490			
Ile	Glu	Leu	Ala	Met	Ala	Leu	Arg	Glu	Thr	Phe	Pro	Gly	Phe	Glu
2495						2500					2505			
Val	Ser	Glu	Leu	Phe	Glu	Thr	Gln	Ser	Ile	Asp	Thr	Leu	Leu	Gly
2510						2515					2520			
Ser	Leu	Glu	Gln	Ala	Pro	Leu	Leu	Ala	Thr	Leu	Thr	Ala	Pro	Pro
2525						2530					2535			
Gln	Gln	Asp	Met	Leu	Gln	Gln	Leu	Lys	Gln	Leu	Leu	Ala	Arg	Thr
2540						2545					2550			
Leu	Lys	Leu	Asp	Ile	Thr	Gln	Ile	Asp	Thr	Ser	Lys	Thr	Leu	Glu
2555						2560					2565			
Ser	Tyr	Gly	Val	Asp	Ser	Ile	Val	Ile	Ile	Glu	Leu	Ala	Asn	Ala
2570						2575					2580			
Leu	Arg	Glu	Arg	Tyr	Pro	Ser	Leu	Asp	Ala	Ser	Gln	Leu	Met	Glu
2585						2590					2595			
Thr	Leu	Ser	Ile	Asp	Arg	Leu	Val	Ala	Gln	Trp	Gln	Ala	Thr	Glu
2600						2605					2610			
Pro	Ala	Val	Pro	Ala	Glu	Pro	Thr	Ala	Glu	Pro	Pro	Val	Ala	Asp
2615						2620					2625			
Glu	Asp	Ala	Ala	Ala	Ile	Ile	Gly	Leu	Ala	Gly	Arg	Phe	Pro	Gly
2630						2635					2640			
Ala	Asp	Thr	Leu	Glu	Glu	Phe	Trp	Asn	Asn	Leu	Arg	Asn	Gly	Gln
2645						2650					2655			
Ser	Ser	Met	Gly	Glu	Val	Pro	Gly	Glu	Arg	Trp	Asp	His	Gln	His
2660						2665					2670			
Tyr	Phe	Asp	Ser	Glu	Arg	Gln	Ala	Pro	Gly	Lys	Thr	Tyr	Ser	Arg
2675						2680					2685			
Trp	Gly	Ala	Phe	Leu	Arg	Asp	Ile	Asp	Gly	Phe	Asp	Ala	Ala	Phe
2690						2695					2700			

Phe Glu Trp Pro Asp Ser Val Ala Leu Glu Ser Asp Pro Gln Ala
 2705 2710 2715
 Arg Ile Phe Leu Glu Gln Ala Tyr Ala Gly Ile Glu Asp Ala Gly
 2720 2725 2730
 Tyr Thr Pro Gly Ser Leu Ser Lys Ser Gln Arg Val Gly Val Phe
 2735 2740 2745
 Val Gly Val Met Asn Gly Tyr Tyr Ser Gly Gly Ala Arg Phe Trp
 2750 2755 2760
 Gln Ile Ala Asn Arg Val Ser Tyr Gln Phe Asp Phe Arg Gly Pro
 2765 2770 2775
 Ser Leu Ala Val Asp Thr Ala Cys Ser Ala Ser Leu Thr Ala Ile
 2780 2785 2790
 His Leu Ala Leu Glu Ser Leu Arg Ser Gly Ser Cys Glu Val Ala
 2795 2800 2805
 Leu Ala Gly Gly Val Asn Leu Leu Val Asp Pro Gln Gln Tyr Leu
 2810 2815 2820
 Asn Leu Ala Gly Ala Ala Met Leu Ser Ala Gly Ala Ser Cys Arg
 2825 2830 2835
 Pro Phe Gly Glu Ala Ala Asp Gly Phe Val Ala Gly Glu Ala Cys
 2840 2845 2850
 Gly Val Val Leu Leu Lys Pro Leu Lys Gln Ala Arg Ala Asp Gly
 2855 2860 2865
 Asp Val Ile His Ala Val Ile Arg Gly Ser Met Ile Asn Ala Gly
 2870 2875 2880
 Gly His Thr Ser Ala Phe Ser Ser Pro Asn Pro Ala Ala Gln Ala
 2885 2890 2895
 Glu Val Val Arg Gln Ala Leu Gln Arg Ala Gly Val Ala Pro Asp
 2900 2905 2910
 Ser Ile Ser Tyr Ile Glu Ala His Gly Thr Gly Thr Val Leu Gly
 2915 2920 2925
 Asp Ala Val Glu Leu Gly Ala Leu Asn Lys Val Phe Asp Lys Arg

2930	2935	2940
Ala Ala Pro Cys Pro Ile Gly Ser Leu Lys Ala Asn Ile Gly His 2945 2950 2955		
Ala Glu Ser Ala Ala Gly Ile Ala Gly Leu Ala Lys Leu Val Leu 2960 2965 2970		
Gln Phe Arg His Gly Glu Leu Val Pro Ser Leu Asn Ala Phe Pro 2975 2980 2985		
Leu Asn Pro Tyr Ile Glu Phe Gly Arg Phe Gln Val Gln Gln Gln 2990 2995 3000		
Pro Ala Pro Trp Pro Arg Arg Gly Ala Gln Pro Arg Arg Ala Gly 3005 3010 3015		
Leu Ser Ala Phe Gly Ala Gly Gly Ser Asn Ala His Leu Val Val 3020 3025 3030		
Glu Glu Ala Pro Ala Met Ala Pro Gly Val Ser Ile Ser Ala Ser 3035 3040 3045		
Ser Pro Ala Leu Ile Val Leu Ser Ala Arg Thr Leu Pro Ala Leu 3050 3055 3060		
Gln Gln Arg Ala Arg Asp Leu Leu Val Trp Met Gln Ala Arg Gln 3065 3070 3075		
Val Asp Asp Val Met Leu Ala Asp Val Ala Tyr Thr Leu His Leu 3080 3085 3090		
Gly Arg Val Ala Met Glu Gln Arg Leu Ala Phe Thr Ala Gly Ser 3095 3100 3105		
Ala Ala Glu Leu Ser Glu Lys Leu Gln Ala Tyr Leu Gly His Ala 3110 3115 3120		
Ile Arg Ala Asp Ile Tyr Leu Ser Glu Asp Thr Pro Gly Lys Pro 3125 3130 3135		
Ala Gly Ala Pro Ile Val Ala Glu Glu Asp Leu Leu Thr Leu Met 3140 3145 3150		
Asp Ala Trp Ile Glu Lys Gly Gln Tyr Gly Arg Leu Leu Glu Tyr 3155 3160 3165		
Trp Thr Lys Gly Gln Pro Ile Asp Trp Asn Lys Leu Tyr Trp Arg		

3170	3175	3180
Lys Leu Tyr Ala Asp Gly Arg 3185	Pro Arg Arg Ile Ser 3190	Leu Pro Thr 3195
Tyr Pro Phe Glu His Arg Arg 3200	Tyr Trp Gln Thr 3205	Pro Val Pro Gly 3210
Glu Arg Ser Leu His Ala Thr 3215	Ala Pro Ala Thr Arg 3220	Glu Thr Val 3225
Ala Val Gly Ala Met Pro Asp 3230	Pro Ala Gly Ala Thr 3235	Val Gln Ala 3240
Arg Leu Cys Ala Leu Cys Gln 3245	Val Leu Leu Gly Lys 3250	Pro Val Thr 3255
Ala Gln Met Asp Phe Phe Ala 3260	Val Gly Gly His Ser 3265	Val Leu Ala 3270
Ile Gln Leu Val Ser Arg Ile 3275	Arg Lys Ser Phe Gly 3280	Val Glu Tyr 3285
Pro Val Ser Ala Leu Phe Glu 3290	Ser Ala Leu Leu Ser 3295	Asp Met Ala 3300
Arg Gln Ile Glu Gln Leu Arg 3305	Val Asn Gly Val Ala 3310	Lys Arg Met 3315
Pro Ala Leu Leu Pro Ala Gly 3320	Arg Val Gly Ala Ile 3325	Pro Ala Thr 3330
Tyr Ala Gln Glu Arg Leu Trp 3335	Leu Val His Glu His 3340	Met Ser Glu 3345
Gln Arg Ser Ser Tyr Asn Ile 3350	Thr Phe Ala Met His 3355	Phe Arg Gly 3360
Val Asp Phe Arg Ala Glu Ala 3365	Met Arg Ala Ala Leu 3370	Asn Ala Leu 3375
Val Val Arg His Glu Val Leu 3380	Arg Thr Arg Phe Leu 3385	Ser Glu Asp 3390
Gly Gln Leu Gln Gln Val Ile 3395	Ala Ala Ser Leu Thr 3400	Leu Glu Val 3405

Pro Val Arg Glu Met Ser Val Glu Glu Val Asp Leu Leu Leu Ala
3410 3415 3420

Ala Ser Thr Arg Glu Thr Phe Asp Leu Arg Gln Gly Pro Leu Phe
3425 3430 3435

Lys Ala Arg Ile Leu Arg Val Ala Ala Asp His His Val Val Leu
3440 3445 3450

Ser Ser Ile His His Ile Ile Ser Asp Gly Trp Ser Leu Gly Val
3455 3460 3465

Phe Asn Arg Asp Leu His Gln Leu Tyr Glu Ala Cys Leu Arg Gly
3470 3475 3480

Thr Pro Pro Thr Leu Pro Thr Leu Ala Val Gln Tyr Ala Asp Tyr
3485 3490 3495

Ala Leu Trp Gln Arg Gln Trp Glu Leu Ala Ala Pro Leu Ser Tyr
3500 3505 3510

Trp Thr Arg Ala Leu Glu Gly Tyr Asp Asp Gly Leu Asp Leu Pro
3515 3520 3525

Tyr Asp Arg Pro Arg Gly Ala Thr Arg Ala Trp Arg Ala Gly Leu
3530 3535 3540

Val Lys His Arg Tyr Pro Pro Gln Leu Ala Gln Gln Leu Ala Ala
3545 3550 3555

Tyr Ser Gln Gln Tyr Gln Ala Thr Leu Phe Met Ser Leu Leu Ala
3560 3565 3570

Gly Leu Ala Leu Val Leu Gly Arg Tyr Ala Asp Arg Lys Asp Val
3575 3580 3585

Cys Ile Gly Ala Thr Val Ser Gly Arg Asp Gln Leu Glu Leu Glu
3590 3595 3600

Glu Leu Ile Gly Phe Phe Ile Asn Ile Leu Pro Leu Arg Val Asp
3605 3610 3615

Leu Ser Gly Asp Pro Cys Leu Glu Glu Val Leu Leu Arg Thr Arg
3620 3625 3630

Gln Val Val Leu Asp Gly Phe Ala His Gln Ser Val Pro Phe Glu
3635 3640 3645

His Val Leu Gln Ala Leu Arg Arg Gln Arg Asp Ser Ser Gln Ile
 3650 3655 3660

Pro Leu Val Pro Val Met Leu Arg His Gln Asn Phe Pro Thr Gln
 3665 3670 3675

Glu Ile Gly Asp Trp Pro Glu Gly Val Arg Leu Thr Gln Met Glu
 3680 3685 3690

Leu Gly Leu Asp Arg Ser Thr Pro Ser Glu Leu Asp Trp Gln Phe
 3695 3700 3705

Tyr Gly Asp Gly Ser Ser Leu Glu Leu Thr Leu Glu Tyr Ala Gln
 3710 3715 3720

Asp Leu Phe Asp Glu Ala Thr Val Arg Arg Met Ile Ala His His
 3725 3730 3735

Gln Gln Ala Leu Glu Ala Met Val Ser Arg Pro Gln Leu Arg Val
 3740 3745 3750

Gly Lys Trp Asp Met Leu Thr Ala Glu Glu Arg Arg Leu Phe Ala
 3755 3760 3765

Ala Leu Asn Ala Thr Gly Thr Pro Arg Glu Trp Pro Ser Leu Ala
 3770 3775 3780

Gln Gln Phe Glu Arg Gln Ala Gln Ala Thr Pro Gln Ala Ile Ala
 3785 3790 3795

Cys Val Ser Asp Gly Gln Ser Trp Ser Tyr Ala Gln Leu Glu Ala
 3800 3805 3810

Arg Ala Asn Gln Leu Ala Gln Ala Leu Arg Gly Gln Gly Ala Gly
 3815 3820 3825

Arg Asp Val Arg Val Ala Val Gln Ser Ala Arg Thr Pro Glu Leu
 3830 3835 3840

Leu Met Ala Leu Leu Ala Ile Phe Lys Ala Gly Ala Cys Tyr Val
 3845 3850 3855

Pro Ile Asp Pro Ala Tyr Pro Ala Ala Tyr Arg Glu Gln Ile Leu
 3860 3865 3870

Ala Glu Val Gln Val Ser Ile Val Leu Glu Gln Asp Glu Leu Ala
 3875 3880 3885

Leu Asp Glu Gln Gly Gln Phe His Asn Pro Arg Trp Arg Glu Gln
 3890 3895 3900
 Ala Pro Thr Pro Leu Gly Leu Arg Glu His Pro Gly Asp Leu Ala
 3905 3910 3915
 Cys Val Met Val Thr Ser Gly Ser Thr Gly Arg Pro Lys Gly Val
 3920 3925 3930
 Met Val Pro Tyr Ala Gln Leu His Asn Trp Leu His Ala Gly Trp
 3935 3940 3945
 Gln Arg Ser Ala Phe Glu Ala Gly Glu Arg Val Leu Gln Lys Thr
 3950 3955 3960
 Ser Ile Ala Phe Ala Val Ser Val Lys Glu Leu Leu Ser Gly Leu
 3965 3970 3975
 Leu Ala Gly Val Glu Gln Val Met Leu Pro Asp Glu Gln Val Lys
 3980 3985 3990
 Asp Ser Leu Ala Leu Ala Arg Ala Ile Glu Gln Trp Gln Val Thr
 3995 4000 4005
 Arg Leu Tyr Leu Val Pro Ser His Leu Gln Ala Leu Leu Asp Ala
 4010 4015 4020
 Thr Gln Gly Arg Asp Gly Leu Leu His Ser Leu Arg His Val Val
 4025 4030 4035
 Thr Ala Gly Glu Ala Leu Pro Ser Ala Val Arg Glu Thr Val Arg
 4040 4045 4050
 Val Arg Leu Pro Gln Val Gln Leu Trp Asn Asn Tyr Gly Cys Thr
 4055 4060 4065
 Glu Leu Asn Asp Ala Thr Tyr His Arg Ser Asp Thr Val Ala Pro
 4070 4075 4080
 Gly Thr Phe Val Pro Ile Gly Ala Pro Ile Ala Asn Thr Glu Val
 4085 4090 4095
 Tyr Val Leu Asp Arg Gln Leu Arg Gln Val Pro Ile Gly Val Met
 4100 4105 4110
 Gly Glu Leu His Val His Ser Val Gly Met Ala Arg Gly Tyr Trp
 4115 4120 4125

Asn Arg Pro Gly Leu Thr Ala Ser Arg Phe Ile Ala His Pro Tyr
 4130 4135 4140

Ser Glu Glu Pro Gly Thr Arg Leu Tyr Lys Thr Gly Asp Met Val
 4145 4150 4155

Arg Arg Leu Ala Asp Gly Thr Leu Glu Tyr Leu Gly Arg Gln Asp
 4160 4165 4170

Phe Glu Val Lys Val Arg Gly His Arg Val Asp Thr Arg Gln Val
 4175 4180 4185

Glu Ala Ala Leu Arg Ala Gln Pro Ala Val Ala Glu Ala Val Val
 4190 4195 4200

Ser Gly His Arg Val Asp Gly Asp Met Gln Leu Val Ala Tyr Val
 4205 4210 4215

Val Ala Arg Glu Gly Gln Ala Pro Ser Ala Gly Glu Leu Lys Gln
 4220 4225 4230

Gln Leu Ser Ala Gln Leu Pro Thr Tyr Met Leu Pro Thr Val Tyr
 4235 4240 4245

Gln Trp Leu Glu Gln Leu Pro Arg Leu Ser Asn Gly Lys Leu Asp
 4250 4255 4260

Arg Leu Ala Leu Pro Ala Pro Gln Val Val His Ala Gln Glu Tyr
 4265 4270 4275

Val Ala Pro Arg Asn Glu Ala Glu Gln Arg Leu Ala Ala Leu Phe
 4280 4285 4290

Ala Glu Val Leu Arg Val Glu Gln Val Gly Ile His Asp Asn Phe
 4295 4300 4305

Phe Ala Leu Gly Gly His Ser Leu Ser Ala Ser Gln Leu Ile Ser
 4310 4315 4320

Arg Ile Arg Gln Ser Phe His Val Asp Leu Pro Leu Ser Arg Ile
 4325 4330 4335

Phe Glu Ala Pro Thr Ile Glu Gly Leu Val Arg Gln Leu Ala Leu
 4340 4345 4350

Pro Ser Glu Gly Gly Val Ala Ser Ile Ala Arg Val Ala Arg Asn

4355	4360	4365
Arg Thr Ile Pro Leu Ser Leu Phe Gln Glu Arg Leu Trp Phe Val		
4370	4375	4380
His Gln His Met Pro Glu Gln Arg Thr Ser Tyr Asn Gly Thr Leu		
4385	4390	4395
Ala Leu Arg Leu Arg Gly Pro Leu Ser Val Glu Ala Met Arg Ala		
4400	4405	4410
Ala Leu Arg Ala Leu Val Leu Arg His Glu Ile Leu Arg Thr Arg		
4415	4420	4425
Phe Val Leu Pro Thr Gly Ala Ser Glu Pro Val Gln Val Ile Asp		
4430	4435	4440
Glu His Ser Asp Phe Gln Leu Ser Val Gln Leu Val Glu Asp Thr		
4445	4450	4455
Glu Ile Ala Ser Leu Met Asp Glu Leu Ala Ser His Ile Tyr Asp		
4460	4465	4470
Leu Ala Asn Gly Pro Leu Phe Ile Ala Cys Leu Leu Gln Leu Asp		
4475	4480	4485
Glu Gln Glu His Val Leu Leu Ile Gly Met His His Leu Ile Tyr		
4490	4495	4500
Asp Ala Trp Ser Gln Phe Thr Val Met Asn Arg Asp Leu Arg Val		
4505	4510	4515
Leu Tyr His Arg His Leu Gly Leu Ala Gly Gly Asp Leu Pro Glu		
4520	4525	4530
Leu Pro Ile Gln Tyr Ala Asp Tyr Ala Ile Trp Gln Arg Ala Gln		
4535	4540	4545
Asn Leu Asp Ala Gln Leu Ala Tyr Trp Gln Ala Met Leu His Asp		
4550	4555	4560
Tyr Asp Asp Gly Leu Glu Leu Pro Tyr Asp Tyr Pro Arg Pro Arg		
4565	4570	4575
Asn Arg Thr Trp His Ala Ala Val Tyr Thr His Thr Tyr Pro Ala		
4580	4585	4590
Glu Leu Val Gln Arg Phe Ala Gly Phe Val Gln Ala His Gln Ser		

4595	4600	4605
Thr Leu Phe Ile Gly Leu Leu Ala Ser Phe Ala Val Val Leu Asn		
4610	4615	4620
Lys Tyr Thr Gly Arg Asp Asp Leu Cys Ile Gly Thr Thr Thr Ala		
4625	4630	4635
Gly Arg Thr His Leu Glu Leu Glu Asn Leu Ile Gly Phe Phe Ile		
4640	4645	4650
Asn Ile Leu Pro Leu Arg Leu Arg Leu Asp Gly Asp Pro Asp Val		
4655	4660	4665
Ala Glu Ile Met Arg Arg Thr Arg Leu Val Ala Met Ser Ala Phe		
4670	4675	4680
Glu Asn Gln Ala Leu Pro Phe Glu His Leu Leu Asn Ala Leu His		
4685	4690	4695
Lys Gln Arg Asp Thr Ser Arg Ile Pro Leu Val Pro Val Val Met		
4700	4705	4710
Arg His Gln Asn Phe Pro Asp Thr Ile Gly Asp Trp Ser Asp Gly		
4715	4720	4725
Ile Arg Thr Glu Val Ile Gln Arg Asp Leu Arg Ala Thr Pro Asn		
4730	4735	4740
Glu Met Asp Leu Gln Phe Phe Gly Asp Gly Thr Gly Leu Ser Val		
4745	4750	4755
Thr Val Glu Tyr Ala Ala Glu Leu Phe Ser Glu Ala Thr Ile Arg		
4760	4765	4770
Arg Leu Ile His His His Gln Leu Val Leu Glu Gln Met Leu Ala		
4775	4780	4785
Ala His Glu Ser Ala Thr Cys Pro Leu Asp Val Ala Asp		
4790	4795	4800

<210> 5
 <211> 45
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1) .. (45)

<223> Acyl-CoA ligase subdomain I

<400> 5

acc tct ggt tcc tcg ggt gag tcc aag ggc atc ctg ctt agc cac
 Thr Ser Gly Ser Ser Gly Glu Ser Lys Gly Ile Leu Leu Ser His
 1 5 10 15

45

<210> 6

<211> 15

<212> PRT

<213> Xanthomonas albilineans

<400> 6

Thr Ser Gly Ser Ser Gly Glu Ser Lys Gly Ile Leu Leu Ser His
 1 5 10 15

<210> 7

<211> 24

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(24)

<223> Acyl-CoA ligase subdomain II

<400> 7

ggt tac ttt cgt acc ggc gac ctg
 Gly Tyr Phe Arg Thr Gly Asp Leu
 1 5

24

<210> 8

<211> 8

<212> PRT

<213> Xanthomonas albilineans

<400> 8

Gly Tyr Phe Arg Thr Gly Asp Leu
 1 5

<210> 9

<211> 51

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(51)

<223> Beta-ketoacyl synthase 1 subdomain I

<400> 9

ggc ccc agc gaa gta atc aac agc gct tgc tcc agc tcg ctg gtg gcg
 Gly Pro Ser Glu Val Ile Asn Ser Ala Cys Ser Ser Ser Leu Val Ala
 1 5 10 15

48

ctg
Leu

51

<210> 10
<211> 17
<212> PRT
<213> Xanthomonas albilineans

<400> 10

Gly Pro Ser Glu Val Ile Asn Ser Ala Cys Ser Ser Ser Leu Val Ala
1 5 10 15

Leu

<210> 11
<211> 30
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(30)
<223> Beta-ketoacyl synthase 1 subdomain II

<400> 11
ggt gaa cta cac ggc act ggt acc agc ctg
Val Glu Leu His Gly Thr Gly Thr Ser Leu
1 5 10

30

<210> 12
<211> 10
<212> PRT
<213> Xanthomonas albilineans

<400> 12

Val Glu Leu His Gly Thr Gly Thr Ser Leu
1 5 10

<210> 13
<211> 30
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(30)
<223> Beta-ketoacyl synthase 1 subdomain III

<400> 13
gcg ctg ggc cat cta gaa gcc gct gca ggc
Ala Leu Gly His Leu Glu Ala Ala Ala Gly

30

1 5 10

<210> 14
 <211> 10
 <212> PRT
 <213> Xanthomonas albilineans

<400> 14

Ala Leu Gly His Leu Glu Ala Ala Ala Gly
 1 5 10

<210> 15
 <211> 51
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(51)
 <223> Beta-ketoacyl synthase 2 subdomain I

<400> 15
 ggg cca agc ctg gcg gtg gat acc gcc tgt tcg gct tcg ctc acc gcg 48
 Gly Pro Ser Leu Ala Val Asp Thr Ala Cys Ser Ala Ser Leu Thr Ala
 1 5 10 15

atc 51
 Ile

<210> 16
 <211> 17
 <212> PRT
 <213> Xanthomonas albilineans

<400> 16
 Gly Pro Ser Leu Ala Val Asp Thr Ala Cys Ser Ala Ser Leu Thr Ala
 1 5 10 15

Ile

<210> 17
 <211> 30
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(30)
 <223> Beta-ketoacyl synthase 2 subdomain II

<400> 17
 atc gag gcg cat ggc acc ggc acc gta cta 30
 Ile Glu Ala His Gly Thr Gly Thr Val Leu

1 5 10

<210> 18
 <211> 10
 <212> PRT
 <213> Xanthomonas albilineans

<400> 18

Ile Glu Ala His Gly Thr Gly Thr Val Leu
 1 5 10

<210> 19
 <211> 30
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(30)
 <223> Beta-ketoacyl synthase 2 subdomain III

<400> 19 30

cgt tct cct cgc cta acc ctg ccg ccc agg
 Arg Ser Pro Arg Leu Thr Leu Pro Pro Arg
 1 5 10

<210> 20
 <211> 10
 <212> PRT
 <213> Xanthomonas albilineans

<400> 20

Arg Ser Pro Arg Leu Thr Leu Pro Pro Arg
 1 5 10

<210> 21
 <211> 93
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(93)
 <223> Beta-ketoacyl reductase domain

<400> 21 48

gtc tac gtc gtg atc ggc ggc gct ggc ggc ttg ggt gaa gta ttg agc
 Val Tyr Val Val Ile Gly Gly Ala Gly Gly Leu Gly Glu Val Leu Ser
 1 5 10 15

gaa cac ttg atc cgc acg tac gac gcg cag ctg atc tgg atc ggg 93
 Glu His Leu Ile Arg Thr Tyr Asp Ala Gln Leu Ile Trp Ile Gly
 20 25 30

<210> 22
 <211> 31
 <212> PRT
 <213> Xanthomonas albilineans

<400> 22

Val Tyr Val Val Ile Gly Gly Ala Gly Gly Leu Gly Glu Val Leu Ser
 1 5 10 15

Glu His Leu Ile Arg Thr Tyr Asp Ala Gln Leu Ile Trp Ile Gly
 20 25 30

<210> 23
 <211> 36
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(36)
 <223> Acyl protein carrier 1 domain

<400> 23
 tgc gaa ctg gcg ctg gat tgc ctc caa tgc gtg cgt
 Cys Glu Leu Ala Leu Asp Ser Leu Gln Cys Val Arg
 1 5 10 36

<210> 24
 <211> 12
 <212> PRT
 <213> Xanthomonas albilineans

<400> 24

Cys Glu Leu Ala Leu Asp Ser Leu Gln Cys Val Arg
 1 5 10

<210> 25
 <211> 36
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(36)
 <223> Acyl carrier protein 2 domain

<400> 25
 gag tac tac ggt gtc gat tgc atc gtg gcg atc gaa
 Glu Tyr Tyr Gly Val Asp Ser Ile Val Ala Ile Glu
 1 5 10 36

<210> 26
 <211> 12
 <212> PRT

<213> Xanthomonas albilineans

<400> 26

Glu Tyr Tyr Gly Val Asp Ser Ile Val Ala Ile Glu
1 5 10

<210> 27

<211> 36

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(36)

<223> Acyl carrier protein 3 domain

<400> 27

gag agc tat ggt gtc gac tcc atc gtc atc atc gaa
Glu Ser Tyr Gly Val Asp Ser Ile Val Ile Ile Glu
1 5 10

36

<210> 28

<211> 12

<212> PRT

<213> Xanthomonas albilineans

<400> 28

Glu Ser Tyr Gly Val Asp Ser Ile Val Ile Ile Glu
1 5 10

<210> 29

<211> 18

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(18)

<223> Adenylation domain subdomain I

<400> 29

tgg agc tat gcg cag ttg
Trp Ser Tyr Ala Gln Leu
1 5

18

<210> 30

<211> 6

<212> PRT

<213> Xanthomonas albilineans

<400> 30

Trp Ser Tyr Ala Gln Leu
1 5

<210> 31
 <211> 33
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(33)
 <223> Adenylation domain subdomain II

<400> 31
 ttc aag gcc ggt gca tgc tat gtg ccg atc gat
 Phe Lys Ala Gly Ala Cys Tyr Val Pro Ile Asp
 1 5 10

33

<210> 32
 <211> 11
 <212> PRT
 <213> Xanthomonas albilineans

<400> 32

Phe Lys Ala Gly Ala Cys Tyr Val Pro Ile Asp
 1 5 10

<210> 33
 <211> 48
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(48)
 <223> Adenylation domain subdomain III

<400> 33
 ctg gcg tgc gtg atg gtg acc tcc ggc tcg acc ggc cgg ccc aag ggc
 Leu Ala Cys Val Met Val Thr Ser Gly Ser Thr Gly Arg Pro Lys Gly
 1 5 10 15

48

<210> 34
 <211> 16
 <212> PRT
 <213> Xanthomonas albilineans

<400> 34

Leu Ala Cys Val Met Val Thr Ser Gly Ser Thr Gly Arg Pro Lys Gly
 1 5 10 15

<210> 35
 <211> 12
 <212> DNA
 <213> Xanthomonas albilineans

<220>

<221> CDS
<222> (1)..(12)
<223> Adenylation domain subdomain IV

<400> 35
ttt gcg gtg tcg
Phe Ala Val Ser
1

12

<210> 36
<211> 4
<212> PRT
<213> Xanthomonas albilineans

<400> 36

Phe Ala Val Ser
1

<210> 37
<211> 21
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(21)
<223> Adenylation domain subdomain V

<400> 37
aac aac tat ggc tgc acg gaa
Asn Asn Tyr Gly Cys Thr Glu
1 5

21

<210> 38
<211> 7
<212> PRT
<213> Xanthomonas albilineans

<400> 38

Asn Asn Tyr Gly Cys Thr Glu
1 5

<210> 39
<211> 45
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(45)
<223> Adenylation domain subdomain VI

<400> 39
ggc gag ctg cac gta cac agc gtg ggg atg gcg cgc ggc tac tgg

45

Gly Glu Leu His Val His Ser Val Gly Met Ala Arg Gly Tyr Trp
 1 5 10 15

<210> 40
 <211> 15
 <212> PRT
 <213> Xanthomonas albilineans

<400> 40

Gly Glu Leu His Val His Ser Val Gly Met Ala Arg Gly Tyr Trp
 1 5 10 15

<210> 41
 <211> 18
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(18)
 <223> Adenylation domain subdomain VII

<400> 41
 tac aag acc ggt gac atg
 Tyr Lys Thr Gly Asp Met
 1 5

18

<210> 42
 <211> 6
 <212> PRT
 <213> Xanthomonas albilineans

<400> 42

Tyr Lys Thr Gly Asp Met
 1 5

<210> 43
 <211> 60
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(60)
 <223> Adenylation domain subdomain VIII

<400> 43
 ggc cga cag gac ttc gag gtc aag gtg cgc ggc cac cgg gtg gat acg
 Gly Arg Gln Asp Phe Glu Val Lys Val Arg Gly His Arg Val Asp Thr
 1 5 10 15

48

cgg cag gtg gag
 Arg Gln Val Glu
 20

60

<210> 44
 <211> 20
 <212> PRT
 <213> Xanthomonas albilineans

<400> 44

Gly Arg Gln Asp Phe Glu Val Lys Val Arg Gly His Arg Val Asp Thr
 1 5 10 15

Arg Gln Val Glu
 20

<210> 45
 <211> 21
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(21)
 <223> Adenylation domain subdomain IX

<400> 45
 atc gcg cac ccg tat agc gag 21
 Ile Ala His Pro Tyr Ser Glu
 1 5

<210> 46
 <211> 7
 <212> PRT
 <213> Xanthomonas albilineans

<400> 46

Ile Ala His Pro Tyr Ser Glu
 1 5

<210> 47
 <211> 18
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(18)
 <223> Adenylation domain subdomain X

<400> 47
 aac ggc aag ttg gac cgg 18
 Asn Gly Lys Leu Asp Arg
 1 5

<210> 48
 <211> 6

<212> PRT
<213> Xanthomonas albilineans

<400> 48

Asn Gly Lys Leu Asp Arg
1 5

<210> 49
<211> 33
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(33)
<223> Peptidyl carrier protein 1 domain

<400> 49
atg gat ttc ttt gcc gtc ggc ggc cat tcg gtg
Met Asp Phe Phe Ala Val Gly Gly His Ser Val
1 5 10 33

<210> 50
<211> 11
<212> PRT
<213> Xanthomonas albilineans

<400> 50
Met Asp Phe Phe Ala Val Gly Gly His Ser Val
1 5 10

<210> 51
<211> 33
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(33)
<223> Peptidyl carrier protein 2 domain

<400> 51
gac aac ttc ttc gcc ttg ggt ggg cac tcg ctg
Asp Asn Phe Phe Ala Leu Gly Gly His Ser Leu
1 5 10 33

<210> 52
<211> 11
<212> PRT
<213> Xanthomonas albilineans

<400> 52
Asp Asn Phe Phe Ala Leu Gly Gly His Ser Leu
1 5 10

<210> 53
 <211> 30
 <212> DNA
 <213> Xanthomonas albilineans

 <220>
 <221> CDS
 <222> (1)..(30)
 <223> Condensation domain 1 subdomain I

<400> 53
 act tat gca cag gag cgc cta tgg ctc gtc
 Thr Tyr Ala Gln Glu Arg Leu Trp Leu Val
 1 5 10

30

<210> 54
 <211> 10
 <212> PRT
 <213> Xanthomonas albilineans

<400> 54
 Thr Tyr Ala Gln Glu Arg Leu Trp Leu Val
 1 5 10

<210> 55
 <211> 27
 <212> DNA
 <213> Xanthomonas albilineans

 <220>
 <221> CDS
 <222> (1)..(27)
 <223> Condensation domain 1 subdomain II

<400> 55
 cgg cac gaa gtg ctg cgc aca cgc ttt
 Arg His Glu Val Leu Arg Thr Arg Phe
 1 5

27

<210> 56
 <211> 9
 <212> PRT
 <213> Xanthomonas albilineans

<400> 56
 Arg His Glu Val Leu Arg Thr Arg Phe
 1 5

<210> 57
 <211> 30
 <212> DNA
 <213> Xanthomonas albilineans
 <220>

<221> CDS
 <222> (1)..(30)
 <223> Condensation domain 1 subdomain III

<400> 57
 atc cac cac atc att tcc gac ggc tgg tcg 30
 ile his his ile ile ser asp gly trp ser
 1 5 10

<210> 58
 <211> 10
 <212> PRT
 <213> Xanthomonas albilineans

<400> 58
 ile his his ile ile ser asp gly trp ser
 1 5 10

<210> 59
 <211> 21
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(21)
 <223> Condensation domain 1 subdomain IV

<400> 59
 tat gcc gac tac gcg ctg tgg 21
 tyr ala asp tyr ala leu trp
 1 5

<210> 60
 <211> 7
 <212> PRT
 <213> Xanthomonas albilineans

<400> 60
 tyr ala asp tyr ala leu trp
 1 5

<210> 61
 <211> 36
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(36)
 <223> Condensation domain 1 subdomain V

<400> 61
 atc ggc ttt ttc atc aat att ttg ccg ctg cgg gtg 36

Ile Gly Phe Phe Ile Asn Ile Leu Pro Leu Arg Val
 1 5 10

<210> 62
 <211> 12
 <212> PRT
 <213> Xanthomonas albilineans

<400> 62

Ile Gly Phe Phe Ile Asn Ile Leu Pro Leu Arg Val
 1 5 10

<210> 63
 <211> 21
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(21)
 <223> Condensation domain 1 subdomain VI

<400> 63
 gcg cac cag tcg gtg ccg ttc
 Ala His Gln Ser Val Pro Phe
 1 5

21

<210> 64
 <211> 7
 <212> PRT
 <213> Xanthomonas albilineans

<400> 64

Ala His Gln Ser Val Pro Phe
 1 5

<210> 65
 <211> 24
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(24)
 <223> Condensation domain 1 subdomain VII

<400> 65
 cgc gac agt agc cag atc ccg ctg
 Arg Asp Ser Ser Gln Ile Pro Leu
 1 5

24

<210> 66
 <211> 8
 <212> PRT

<213> Xanthomonas albilineans

<400> 66

Arg Asp Ser Ser Gln Ile Pro Leu
1 5

<210> 67

<211> 30

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(30)

<223> Condensation domain 2 subdomain I

<400> 67

tcg ctg ttc cag gaa cgc ctg tgg ttc gtg
Ser Leu Phe Gln Glu Arg Leu Trp Phe Val
1 5 10

30

<210> 68

<211> 10

<212> PRT

<213> Xanthomonas albilineans

<400> 68

Ser Leu Phe Gln Glu Arg Leu Trp Phe Val
1 5 10

<210> 69

<211> 27

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(27)

<223> Condensation domain 2 subdomain II

<400> 69

cgc cac gaa atc ttg cgt acc cgc ttc
Arg His Glu Ile Leu Arg Thr Arg Phe
1 5

27

<210> 70

<211> 9

<212> PRT

<213> Xanthomonas albilineans

<400> 70

Arg His Glu Ile Leu Arg Thr Arg Phe
1 5

<210> 71
<211> 30
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(30)
<223> Condensation domain 2 subdomain III

<400> 71
atg cat cac ctt atc tac gac gct tgg tcg
Met His His Leu Ile Tyr Asp Ala Trp Ser
1 5 10

30

<210> 72
<211> 10
<212> PRT
<213> Xanthomonas albilineans

<400> 72
Met His His Leu Ile Tyr Asp Ala Trp Ser
1 5 10

<210> 73
<211> 21
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(21)
<223> Condensation domain 2 subdomain IV

<400> 73
tat gcc gac tat gcg atc tgg
Tyr Ala Asp Tyr Ala Ile Trp
1 5

21

<210> 74
<211> 7
<212> PRT
<213> Xanthomonas albilineans

<400> 74
Tyr Ala Asp Tyr Ala Ile Trp
1 5

<210> 75
<211> 33
<212> DNA
<213> Xanthomonas albilineans

<220>

<221> CDS
 <222> (1)..(33)
 <223> Condensation domain 2 subdomain V

<400> 75
 atc ggt ttc ttc atc aac atc ttg cct ttg cgc
 ile gly phe phe ile asn ile leu pro leu arg
 1 5 10

33

<210> 76
 <211> 11
 <212> PRT
 <213> Xanthomonas albilineans

<400> 76
 ile gly phe phe ile asn ile leu pro leu arg
 1 5 10

<210> 77
 <211> 21
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(21)
 <223> Condensation domain 2 subdomain VI

<400> 77
 aac cag gcg cta ccg ttc gag
 asn gln ala leu pro phe glu
 1 5

21

<210> 78
 <211> 7
 <212> PRT
 <213> Xanthomonas albilineans

<400> 78
 asn gln ala leu pro phe glu
 1 5

<210> 79
 <211> 24
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(24)
 <223> Condensation domain 2 subdomain VII

<400> 79
 cgt gac acc agc cgg att ccg cta
 arg asp thr ser arg ile pro leu

24

1

5

<210> 80
 <211> 8
 <212> PRT
 <213> Xanthomonas albilineans

<400> 80

Arg Asp Thr Ser Arg Ile Pro Leu
 1 5

<210> 81
 <211> 242
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> promoter
 <222> (1)..(242)
 <223> Bidirectional xab8 promoter

<400> 81
 catcacgccca cctccagcag ggtgtcatat acggccagcg gatgctgcag gttttccac: 60
 ggcaggggcca ctggtgtcg taagggaagc ggtgccttga gcgccggtgc ggacagtata 120
 acgacacggt ccttggccaa gcgcactgtc ggcacggcct tgctgatgcc gcccatgtag 180
 ccgcgcgcct ggatctcgcg tagtagcacc acgctggccg ggatccatcg agggcgcgct 240
 tg 242

<210> 82
 <211> 1200
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (149)..(982)
 <223>

<400> 82
 gctctgtccc gcgtcgtcca tcgccattgc gcccctcccc gacccaagc atcgaccaa- 60
 ggaccgaatg cggcggttag gcgcgactct gcgacactag cgcaatgtta tcgtcgacat 120
 tgacgcccac agccctcagc gcaacgca atg ccc aat gcc gta ccg atg cag 172
 Met Pro Asn Ala Val Pro Met Gln
 1 5
 ggc gcg cgg gga ctc ccg cag ccg caa gcg atg aac cca ggg ttg ccg 220
 Gly Ala Arg Gly Leu Pro Gln Pro Gln Ala Met Asn Pro Gly Leu Pro
 10 15 20
 agc gtc ggc ggc ttg agc gca ggc cag cca ttg cag ttg tcg tta gca 268
 Ser Val Gly Gly Leu Ser Ala Gly Gln Pro Leu Gln Leu Ser Leu Ala
 25 30 35 40

ccg gaa ctg cag gca gcc gcg cgc agt gcc cac cgc cat ctg ctc gac Pro Glu Leu Gln Ala Ala Ala Arg Ser Ala His Arg His Leu Leu Asp 45 50 55	316
gac ggc acg gcg ctt tac ctg ctg gcg ttc gat acc gcg caa ttc gac Asp Gly Thr Ala Leu Tyr Leu Leu Ala Phe Asp Thr Ala Gln Phe Asp 60 65 70	364
ccg ggg gct ttc gcg gca atg gca atc gcc cgc ccg gac agc atc gcc Pro Gly Ala Phe Ala Ala Met Ala Ile Ala Arg Pro Asp Ser Ile Ala 75 80 85	412
cgc agc gtg cgc aag cgt cag gcc gag ttc ctg ttc ggc cgt ctg gcc Arg Ser Val Arg Lys Arg Gln Ala Glu Phe Leu Phe Gly Arg Leu Ala 90 95 100	460
gcg cga ctg gcg ctg caa gag gtg ctg gga cct gcg caa gcg cag gca Ala Arg Leu Ala Leu Gln Glu Val Leu Gly Pro Ala Gln Ala Gln Ala 105 110 115 120	508
gat att gca atc ggc gcg acg cgc gcg ccc tgc tgg cct gcc ggc agc Asp Ile Ala Ile Gly Ala Thr Arg Ala Pro Cys Trp Pro Ala Gly Ser 125 130 135	556
ctg ggc agc att tcc cat tgc gag gac tac gcg gcc gcc atc gcc atg Leu Gly Ser Ile Ser His Cys Glu Asp Tyr Ala Ala Ala Ile Ala Met 140 145 150	604
gcg gcc ggc acc cgc cac gcc gtg ggc atc gat ctg gaa cga cca atc Ala Ala Gly Thr Arg His Gly Val Gly Ile Asp Leu Glu Arg Pro Ile 155 160 165	652
aca ccc gcg gcg cgc gcg gcg ttg ctg agc atc gca atc gat gcc gac Thr Pro Ala Ala Arg Ala Ala Leu Leu Ser Ile Ala Ile Asp Ala Asp 170 175 180	700
gaa gcc gct cgt ctg gca aag gcg gca gac gcg cag tgg ccg caa gac Glu Ala Ala Arg Leu Ala Lys Ala Ala Asp Ala Gln Trp Pro Gln Asp 185 190 195 200	748
ctg ctg ctg acc gca cta ttt tcg gcc aag gaa agc ctg ttc aaa gcc Leu Leu Leu Thr Ala Leu Phe Ser Ala Lys Glu Ser Leu Phe Lys Ala 205 210 215	796
gcc tac agc gcg gtc gga cgc tac ttc gac ttc agc gcg gca cgc ctg Ala Tyr Ser Ala Val Gly Arg Tyr Phe Asp Phe Ser Ala Ala Arg Leu 220 225 230	844
tgc ggc atc gac ctg gca cgg caa tgc ctg cat ctg cgc ctg acc gag Cys Gly Ile Asp Leu Ala Arg Gln Cys Leu His Leu Arg Leu Thr Glu 235 240 245	892
aca ctc tgc gcg caa ttc gtg gcc ggg caa gtg tgc gag gtc ggc ttc Thr Leu Cys Ala Gln Phe Val Ala Gly Gln Val Cys Glu Val Gly Phe 250 255 260	940
gcg cgc cta cca ccg gac ctg gtg ctc acc cac tac gcc tgg Ala Arg Leu Pro Pro Asp Leu Val Leu Thr His Tyr Ala Trp 265 270 275	982
tgagcacgcg gacagtcgaa cccgccaacg ccaacggcac tcaagacgtg gcgtgcgccg	1042
cgctcggtcgt gaagctctcc ccgcagccgc actcggcggtt ggcattggga ttgcggaaca	1102

cgaaggtctc acccaagccc tgcttggcga agtcgatttc ggtgccatcg accaactgca 1162

gactggcggc atcgacataa atccgcactc cgtcctgc 1200

<210> 83

<211> 278

<212> PRT

<213> Xanthomonas albilineans

<400> 83

Met Pro Asn Ala Val Pro Met Gln Gly Ala Arg Gly Leu Pro Gln Pro
1 5 10 15

Gln Ala Met Asn Pro Gly Leu Pro Ser Val Gly Gly Leu Ser Ala Gly
20 25 30

Gln Pro Leu Gln Leu Ser Leu Ala Pro Glu Leu Gln Ala Ala Ala Arg
35 40 45

Ser Ala His Arg His Leu Leu Asp Asp Gly Thr Ala Leu Tyr Leu Leu
50 55 60

Ala Phe Asp Thr Ala Gln Phe Asp Pro Gly Ala Phe Ala Ala Met Ala
65 70 75 80

Ile Ala Arg Pro Asp Ser Ile Ala Arg Ser Val Arg Lys Arg Gln Ala
85 90 95

Glu Phe Leu Phe Gly Arg Leu Ala Ala Arg Leu Ala Leu Gln Glu Val
100 105 110

Leu Gly Pro Ala Gln Ala Gln Ala Asp Ile Ala Ile Gly Ala Thr Arg
115 120 125

Ala Pro Cys Trp Pro Ala Gly Ser Leu Gly Ser Ile Ser His Cys Glu
130 135 140

Asp Tyr Ala Ala Ala Ile Ala Met Ala Ala Gly Thr Arg His Gly Val
145 150 155 160

Gly Ile Asp Leu Glu Arg Pro Ile Thr Pro Ala Ala Arg Ala Ala Leu
165 170 175

Leu Ser Ile Ala Ile Asp Ala Asp Glu Ala Ala Arg Leu Ala Lys Ala
180 185 190

Ala Asp Ala Gln Trp Pro Gln Asp Leu Leu Leu Thr Ala Leu Phe Ser
195 200 205

Ala Lys Glu Ser Leu Phe Lys Ala Ala Tyr Ser Ala Val Gly Arg Tyr
210 215 220

Phe Asp Phe Ser Ala Ala Arg Leu Cys Gly Ile Asp Leu Ala Arg Gln
225 230 235 240

Cys Leu His Leu Arg Leu Thr Glu Thr Leu Cys Ala Gln Phe Val Ala
245 250 255

Gly Gln Val Cys Glu Val Gly Phe Ala Arg Leu Pro Pro Asp Leu Val
260 265 270

Leu Thr His Tyr Ala Trp
275

<210> 84
<211> 837
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(837)
<223>

<400> 84
atg ccc aat gcc gta ccg atg cag gcc gcg cgg gga ctc ccg cag ccg 48
Met Pro Asn Ala Val Pro Met Gln Gly Ala Arg Gly Leu Pro Gln Pro
1 5 10 15
caa gcg atg aac cca ggg ttg ccg agc gtc gcc gcc ttg agc gca gcc 96
Gln Ala Met Asn Pro Gly Leu Pro Ser Val Gly Gly Leu Ser Ala Gly
20 25 30
cag cca ttg cag ttg tcg tta gca ccg gaa ctg cag gca gcc gcg cgc 144
Gln Pro Leu Gln Leu Ser Leu Ala Pro Glu Leu Gln Ala Ala Ala Arg
35 40 45
agt gcc cac cgc cat ctg ctc gac gac gcc acg gcg ctt tac ctg ctg 192
Ser Ala His Arg His Leu Leu Asp Asp Gly Thr Ala Leu Tyr Leu Leu
50 55 60
gcg ttc gat acc gcg caa ttc gac ccg ggg gct ttc gcg gca atg gca 240
Ala Phe Asp Thr Ala Gln Phe Asp Pro Gly Ala Phe Ala Ala Met Ala
65 70 75 80
atc gcc cgc ccg gac agc atc gcc cgc agc gtg cgc aag cgt cag gcc 288
Ile Ala Arg Pro Asp Ser Ile Ala Arg Ser Val Arg Lys Arg Gln Ala
85 90 95
gag ttc ctg ttc gcc cgt ctg gcc gcg cga ctg gcg ctg caa gag gtg 336
Glu Phe Leu Phe Gly Arg Leu Ala Ala Arg Leu Ala Leu Gln Glu Val
100 105 110
ctg gga cct gcg caa gcg cag gca gat att gca atc gcc gcg acg cgc 384
Leu Gly Pro Ala Gln Ala Gln Ala Asp Ile Ala Ile Gly Ala Thr Arg
115 120 125

gcg ccc tgc tgg cct gcc ggc agc ctg ggc agc att tcc cat tgc gag 432
 Ala Pro Cys Trp Pro Ala Gly Ser Leu Gly Ser Ile Ser His Cys Glu
 130 135 140

gac tac gcg gcc gcc atc gcc atg gcg gcc ggc acc cgc cac ggc gtg 480
 Asp Tyr Ala Ala Ala Ile Ala Met Ala Ala Gly Thr Arg His Gly Val
 145 150 155 160

ggc atc gat ctg gaa cga cca atc aca ccc gcg gcg cgc gcg gcg ttg 528
 Gly Ile Asp Leu Glu Arg Pro Ile Thr Pro Ala Ala Arg Ala Ala Leu
 165 170 175

ctg agc atc gca atc gat gcc gac gaa gcc gct cgt ctg gca aag gcg 576
 Leu Ser Ile Ala Ile Asp Ala Asp Glu Ala Ala Arg Leu Ala Lys Ala
 180 185 190

gca gac gcg cag tgg ccg caa gac ctg ctg ctg acc gca cta ttt tcg 624
 Ala Asp Ala Gln Trp Pro Gln Asp Leu Leu Leu Thr Ala Leu Phe Ser
 195 200 205

gcc aag gaa agc ctg ttc aaa gcc gcc tac agc gcg gtc gga cgc tac 672
 Ala Lys Glu Ser Leu Phe Lys Ala Ala Tyr Ser Ala Val Gly Arg Tyr
 210 215 220

ttc gac ttc agc gcg gca cgc ctg tgc ggc atc gac ctg gca cgg caa 720
 Phe Asp Phe Ser Ala Ala Arg Leu Cys Gly Ile Asp Leu Ala Arg Gln
 225 230 235 240

tgc ctg cat ctg cgc ctg acc gag aca ctc tgc gcg caa ttc gtg gcc 768
 Cys Leu His Leu Arg Leu Thr Glu Thr Leu Cys Ala Gln Phe Val Ala
 245 250 255

ggg caa gtg tgc gag gtc ggc ttc gcg cgc cta cca ccg gac ctg gtg 816
 Gly Gln Val Cys Glu Val Gly Phe Ala Arg Leu Pro Pro Asp Leu Val
 260 265 270

ctc acc cac tac gcc tgg tga 837
 Leu Thr His Tyr Ala Trp
 275

<210> 85
 <211> 278
 <212> PRT
 <213> Xanthomonas albilineans

<400> 85

Met Pro Asn Ala Val Pro Met Gln Gly Ala Arg Gly Leu Pro Gln Pro
 1 5 10 15

Gln Ala Met Asn Pro Gly Leu Pro Ser Val Gly Gly Leu Ser Ala Gly
 20 25 30

Gln Pro Leu Gln Leu Ser Leu Ala Pro Glu Leu Gln Ala Ala Arg
 35 40 45

Ser Ala His Arg His Leu Leu Asp Asp Gly Thr Ala Leu Tyr Leu Leu
 50 55 60

Ala Phe Asp Thr Ala Gln Phe Asp Pro Gly Ala Phe Ala Ala Met Ala
65 70 75 80

Ile Ala Arg Pro Asp Ser Ile Ala Arg Ser Val Arg Lys Arg Gln Ala
85 90 95

Glu Phe Leu Phe Gly Arg Leu Ala Ala Arg Leu Ala Leu Gln Glu Val
100 105 110

Leu Gly Pro Ala Gln Ala Gln Ala Asp Ile Ala Ile Gly Ala Thr Arg
115 120 125

Ala Pro Cys Trp Pro Ala Gly Ser Leu Gly Ser Ile Ser His Cys Glu
130 135 140

Asp Tyr Ala Ala Ala Ile Ala Met Ala Ala Gly Thr Arg His Gly Val
145 150 155 160

Gly Ile Asp Leu Glu Arg Pro Ile Thr Pro Ala Ala Arg Ala Ala Leu
165 170 175

Leu Ser Ile Ala Ile Asp Ala Asp Glu Ala Ala Arg Leu Ala Lys Ala
180 185 190

Ala Asp Ala Gln Trp Pro Gln Asp Leu Leu Leu Thr Ala Leu Phe Ser
195 200 205

Ala Lys Glu Ser Leu Phe Lys Ala Ala Tyr Ser Ala Val Gly Arg Tyr
210 215 220

Asp Phe Ser Ala Ala Arg Leu Cys Gly Ile Asp Leu Ala Arg Gln
225 230 235 240

Cys Leu His Leu Arg Leu Thr Glu Thr Leu Cys Ala Gln Phe Val Ala
245 250 255

Gly Gln Val Cys Glu Val Gly Phe Ala Arg Leu Pro Pro Asp Leu Val
260 265 270

Leu Thr His Tyr Ala Trp
275

<210> 86
<211> 180
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS

<222> (1)..(168)

<223>

<400> 86

ggc gtg ggc atc gat ctg gaa cga cca atc aca ccc gcg gcg cgc gcg 48
 Gly Val Gly Ile Asp Leu Glu Arg Pro Ile Thr Pro Ala Ala Arg Ala
 1 5 10 15

gcg ttg ctg agc atc gca atc gat gcc gac gaa gcc gct cgt ctg gca 96
 Ala Leu Leu Ser Ile Ala Ile Asp Ala Asp Glu Ala Ala Arg Leu Ala
 20 25 30

aag gcg gca gac gcg cag tgg ccg caa gac ctg ctg ctg acc gca cta 144
 Lys Ala Ala Asp Ala Gln Trp Pro Gln Asp Leu Leu Leu Thr Ala Leu
 35 40 45

ttt tct gcc aag gaa agc ctg ttc aaagccgcct ac 180
 Phe Ser Ala Lys Glu Ser Leu Phe
 50 55

<210> 87

<211> 56

<212> PRT

<213> Xanthomonas albilineans

<400> 87

Gly Val Gly Ile Asp Leu Glu Arg Pro Ile Thr Pro Ala Ala Arg Ala
 1 5 10 15

Ala Leu Leu Ser Ile Ala Ile Asp Ala Asp Glu Ala Ala Arg Leu Ala
 20 25 30

Lys Ala Ala Asp Ala Gln Trp Pro Gln Asp Leu Leu Leu Thr Ala Leu
 35 40 45

Phe Ser Ala Lys Glu Ser Leu Phe
 50 55

<210> 88

<211> 27

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(27)

<223>

<400> 88

ggc gtg ggc atc gat ctg gaa cga cca 27
 Gly Val Gly Ile Asp Leu Glu Arg Pro
 1 5

<210> 89

<211> 9
 <212> PRT
 <213> Xanthomonas albilineans

<400> 89

Gly Val Gly Ile Asp Leu Glu Arg Pro
 1 5

<210> 90
 <211> 117
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(117)
 <223>

<400> 90
 atc aca ccc gcg gcg cgc gcg gcg ttg ctg agc atc gca atc gat gcc 48
 Ile Thr Pro Ala Ala Arg Ala Ala Leu Leu Ser Ile Ala Ile Asp Ala
 1 5 10 15

gac gaa gcc gct cgt ctg gca aag gcg gca gac gcg cag tgg ccg caa 96
 Asp Glu Ala Ala Arg Leu Ala Lys Ala Ala Asp Ala Gln Trp Pro Gln
 20 25 30

gac ctg ctg ctg acc gca cta 117
 Asp Leu Leu Leu Thr Ala Leu
 35

<210> 91
 <211> 39
 <212> PRT
 <213> Xanthomonas albilineans

<400> 91

Ile Thr Pro Ala Ala Arg Ala Ala Leu Leu Ser Ile Ala Ile Asp Ala
 1 5 10 15

Asp Glu Ala Ala Arg Leu Ala Lys Ala Ala Asp Ala Gln Trp Pro Gln
 20 25 30

Asp Leu Leu Leu Thr Ala Leu
 35

<210> 92
 <211> 36
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(36)
 <223>

<400> 92

ttt tcg gcc aag gaa agc ctg ttc aaa gcc gcc tac
 Phe Ser Ala Lys Glu Ser Leu Phe Lys Ala Ala Tyr
 1 5 10

36

<210> 93

<211> 12

<212> PRT

<213> Xanthomonas albilineans

<400> 93

Phe Ser Ala Lys Glu Ser Leu Phe Lys Ala Ala Tyr
 1 5 10

<210> 94

<211> 1515

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (46)..(1071)

<223>

<400> 94

caaaagccgg ccgccgtcac ccgttcacgc atagcgaggg caatc atg gat tca gcg
 Met Asp Ser Ala
 1

57

tta cct aca tct gca ttt acc ttc gat ctc ttt tac acc acg gtt aac
 Leu Pro Thr Ser Ala Phe Thr Phe Asp Leu Phe Tyr Thr Thr Val Asn
 5 10 15 20

105

gcc tac tat cgc act gcc gca gtc aag gcg gcg atc gaa ctg ggg cta
 Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile Glu Leu Gly Leu
 25 30 35

153

ttc gat gtg gtg ggg cag cag ggc cga act ccc gca gcc atc gcc gag
 Phe Asp Val Val Gly Gln Gln Gly Arg Thr Pro Ala Ala Ile Ala Glu
 40 45 50

201

gcc tgc cag gcg tcg ccg cgc ggc att cgc atc ctt tgc tat tac cta
 Ala Cys Gln Ala Ser Pro Arg Gly Ile Arg Ile Leu Cys Tyr Tyr Leu
 55 60 65

249

gta tcg atc ggt ttt cta cgc cgc aac ggt ggc ctg ttc tac ata gat
 Val Ser Ile Gly Phe Leu Arg Arg Asn Gly Gly Leu Phe Tyr Ile Asp
 70 75 80

297

cgc aac atg gcc atg tac ctg gat cgt agt tcg ccc ggc tac ctg ggt
 Arg Asn Met Ala Met Tyr Leu Asp Arg Ser Ser Pro Gly Tyr Leu Gly
 85 90 95 100

345

ggc agc atc aag ttc ctg ctc tcg ccc tac atc atg agc gcc ttc acc
 Gly Ser Ile Lys Phe Leu Leu Ser Pro Tyr Ile Met Ser Ala Phe Thr
 105 110 115

393

gat ctg acc gcc gta gtc agg acc ggc aag atc aac ctg gcg cag gac Asp Leu Thr Ala Val Val Arg Thr Gly Lys Ile Asn Leu Ala Gln Asp 120 125 130	441
ggc gtg gtg gca ccg gat cac ccg cag tgg gtg gaa ttt gca cgc gcg Gly Val Val Ala Pro Asp His Pro Gln Trp Val Glu Phe Ala Arg Ala 135 140 145	489
atg gca ccg atg atg gcg ctg ccc tcg gcg ttg atc gcc aat atg gtg Met Ala Pro Met Met Ala Leu Pro Ser Ala Leu Ile Ala Asn Met Val 150 155 160	537
tcg ttg ccc gct gat cgg ccg att cgt gtg ctg gac gtg gca gcc ggc Ser Leu Pro Ala Asp Arg Pro Ile Arg Val Leu Asp Val Ala Ala Gly 165 170 175 180	585
cac ggc ctg ttc ggc atc gcc ttc gcg cag cgc ttc cgc cag gct gaa His Gly Leu Phe Gly Ile Ala Phe Ala Gln Arg Phe Arg Gln Ala Glu 185 190 195	633
gtg agc ttc ctg gac tgg gac aac gtg cta gac gta gca cgc gaa aac Val Ser Phe Leu Asp Trp Asp Asn Val Leu Asp Val Ala Arg Glu Asn 200 205 210	681
gcc cag gcg gcc aaa gtg gcc gag cga gcg cgt ttc ctg ccc ggc aac Ala Gln Ala Ala Lys Val Ala Glu Arg Ala Arg Phe Leu Pro Gly Asn 215 220 225	729
gca ttc gac ctc gat tac ggc agc ggc tac gac gtg atc ttg ttg acc Ala Phe Asp Leu Asp Tyr Gly Ser Gly Tyr Asp Val Ile Leu Leu Thr 230 235 240	777
aac ttc ctg cac cat ttc gat gag gtc gat ggc gag cgc atc ttg gct Asn Phe Leu His His Phe Asp Glu Val Asp Gly Glu Arg Ile Leu Ala 245 250 255 260	825
aag acg cgc gat gcg ctg aac gac gac ggc atg gtg atc act ttc gaa Lys Thr Arg Asp Ala Leu Asn Asp Asp Gly Met Val Ile Thr Phe Glu 265 270 275	873
ttc atc gcc gac gaa gag cgt tcc tca ccg ccg ctg gcc gcc acc ttc Phe Ile Ala Asp Glu Glu Arg Ser Ser Pro Pro Leu Ala Ala Thr Phe 280 285 290	921
agc atg atg atg ctg ggc acc acc ccg gcg ggc gag tcc tac acc tat Ser Met Met Met Leu Gly Thr Thr Pro Ala Gly Glu Ser Tyr Thr Tyr 295 300 305	969
agc gat ctg gaa agg atg ttt cgg cat gcc ggc ttc ggc cac gtg gaa Ser Asp Leu Glu Arg Met Phe Arg His Ala Gly Phe Gly His Val Glu 310 315 320	1017
cta aaa tcg ata ccg ccg gcc ttg ctg aaa gtg gtg gtt tcc cgc aag Leu Lys Ser Ile Pro Pro Ala Leu Leu Lys Val Val Ser Arg Lys 325 330 335 340	1065
agg gcc ccataatgat cgaatcggcg acatcccctg tggcgaaaac cgagcgcac Arg Ala	1121
tggtgcaccg agctggacct ggatgcactc aacgccatgt cggccaacac gatgcaggcc	1181
ctgctcggta tacgcatgat cgagatcggc tcggactatc tggtctcctg catgtcgggt	1241

gactggcggt gccaccagcc ctatggggta ttgcatggcg gcgcatcggt caccctggcc 1301
 gaggtaccg gcagcatggc ggctccatg tgcgtgccg ccggccaacg ttgcgttggc 1361
 ctagacatca atgccaacca catcgcgagc atctccagt gccaagtaca gtgcatcgcy 1421
 cggccgctgc acataggggc cttgaccag gtatggcaga tgcgcatcta tgacgaaggt 1481
 gaccgcacga tctgcgtgtc gcgcctgacc atgg 1515

<210> 95
 <211> 342
 <212> PRT
 <213> Xanthomonas albilineans
 <400> 95

Met Asp Ser Ala Leu Pro Thr Ser Ala Phe Thr Phe Asp Leu Phe Tyr
 1 5 10 15

Thr Thr Val Asn Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile
 20 25 30

Glu Leu Gly Leu Phe Asp Val Val Gly Gln Gln Gly Arg Thr Pro Ala
 35 40 45

Ala Ile Ala Glu Ala Cys Gln Ala Ser Pro Arg Gly Ile Arg Ile Leu
 50 55 60

Cys Tyr Tyr Leu Val Ser Ile Gly Phe Leu Arg Arg Asn Gly Gly Leu
 65 70 75 80

Phe Tyr Ile Asp Arg Asn Met Ala Met Tyr Leu Asp Arg Ser Ser Pro
 85 90 95

Gly Tyr Leu Gly Gly Ser Ile Lys Phe Leu Leu Ser Pro Tyr Ile Met
 100 105 110

Ser Ala Phe Thr Asp Leu Thr Ala Val Val Arg Thr Gly Lys Ile Asn
 115 120 125

Leu Ala Gln Asp Gly Val Val Ala Pro Asp His Pro Gln Trp Val Glu
 130 135 140

Phe Ala Arg Ala Met Ala Pro Met Met Ala Leu Pro Ser Ala Leu Ile
 145 150 155 160

Ala Asn Met Val Ser Leu Pro Ala Asp Arg Pro Ile Arg Val Leu Asp
 165 170 175

Val Ala Ala Gly His Gly Leu Phe Gly Ile Ala Phe Ala Gln Arg Phe

180	185	190
Arg Gln Ala Glu Val Ser Phe Leu Asp Trp Asp Asn Val Leu Asp Val		
195	200	205
Ala Arg Glu Asn Ala Gln Ala Ala Lys Val Ala Glu Arg Ala Arg Phe		
210	215	220
Leu Pro Gly Asn Ala Phe Asp Leu Asp Tyr Gly Ser Gly Tyr Asp Val		
225	230	235
Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Val Asp Gly Glu		
245	250	255
Arg Ile Leu Ala Lys Thr Arg Asp Ala Leu Asn Asp Asp Gly Met Val		
260	265	270
Ile Thr Phe Glu Phe Ile Ala Asp Glu Glu Arg Ser Ser Pro Pro Leu		
275	280	285
Ala Ala Thr Phe Ser Met Met Met Leu Gly Thr Thr Pro Ala Gly Glu		
290	295	300
Ser Tyr Thr Tyr Ser Asp Leu Glu Arg Met Phe Arg His Ala Gly Phe		
305	310	315
Gly His Val Glu Leu Lys Ser Ile Pro Pro Ala Leu Leu Lys Val Val		
325	330	335
Val Ser Arg Lys Arg Ala		
340		

<210> 96
 <211> 1032
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(1029)
 <223>

<400> 96	
atg gat tca gcg tta cct aca tct gca ttt acc ttc gat ctc ttt tac	48
Met Asp Ser Ala Leu Pro Thr Ser Ala Phe Thr Phe Asp Leu Phe Tyr	
1 5 10 15	
acc acg gtt aac gcc tac tat cgc act gcc gca gtc aag gcg gcg atc	96
Thr Thr Val Asn Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile	
20 25 30	
gaa ctg ggg cta ttc gat gtg gtg ggg cag cag ggc cga act ccc gca	144

Glu	Leu	Gly	Leu	Phe	Asp	Val	Val	Gly	Gln	Gln	Gly	Arg	Thr	Pro	Ala		
	35						40					45					
gcc	atc	gcc	gag	gcc	tgc	cag	gcg	tcg	ccg	cgc	ggc	att	cgc	atc	ctt	192	
Ala	Ile	Ala	Glu	Ala	Cys	Gln	Ala	Ser	Pro	Arg	Gly	Ile	Arg	Ile	Leu		
	50					55				60							
tgc	tat	tac	cta	gta	tcg	atc	ggg	ttt	cta	cgc	cgc	aac	ggg	ggc	ctg	240	
Cys	Tyr	Tyr	Leu	Val	Ser	Ile	Gly	Phe	Leu	Arg	Arg	Asn	Gly	Gly	Leu		
65					70				75					80			
ttc	tac	ata	gat	cgc	aac	atg	gcc	atg	tac	ctg	gat	cgt	agt	tcg	ccc	288	
Phe	Tyr	Ile	Asp	Arg	Asn	Met	Ala	Met	Tyr	Leu	Asp	Arg	Ser	Ser	Pro		
				85					90					95			
ggc	tac	ctg	ggg	ggc	agc	atc	aag	ttc	ctg	ctc	tcg	ccc	tac	atc	atg	336	
Gly	Tyr	Leu	Gly	Gly	Ser	Ile	Lys	Phe	Leu	Leu	Ser	Pro	Tyr	Ile	Met		
		100					105						110				
agc	gcc	ttc	acc	gat	ctg	acc	gcc	gta	gtc	agg	acc	ggc	aag	atc	aac	384	
Ser	Ala	Phe	Thr	Asp	Leu	Thr	Ala	Val	Val	Arg	Thr	Gly	Lys	Ile	Asn		
		115					120					125					
ctg	gcg	cag	gac	ggc	gtg	gtg	gca	ccg	gat	cac	ccg	cag	tgg	gtg	gaa	432	
Leu	Ala	Gln	Asp	Gly	Val	Val	Ala	Pro	Asp	His	Pro	Gln	Trp	Val	Glu		
	130				135						140						
ttt	gca	cgc	gcg	atg	gca	ccg	atg	atg	gcg	ctg	ccc	tcg	gcg	ttg	atc	480	
Phe	Ala	Arg	Ala	Met	Ala	Pro	Met	Met	Ala	Leu	Pro	Ser	Ala	Leu	Ile		
145					150				155					160			
gcc	aat	atg	gtg	tcg	ttg	ccc	gct	gat	cgg	ccg	att	cgt	gtg	ctg	gac	528	
Ala	Asn	Met	Val	Ser	Leu	Pro	Ala	Asp	Arg	Pro	Ile	Arg	Val	Leu	Asp		
				165				170					175				
gtg	gca	gcc	ggc	cac	ggc	ctg	ttc	ggc	atc	gcc	ttc	gcg	cag	cgc	ttc	576	
Val	Ala	Ala	Gly	His	Gly	Leu	Phe	Gly	Ile	Ala	Phe	Ala	Gln	Arg	Phe		
		180				185						190					
cgc	cag	gct	gaa	gtg	agc	ttc	ctg	gac	tgg	gac	aac	gtg	cta	gac	gta	624	
Arg	Gln	Ala	Glu	Val	Ser	Phe	Leu	Asp	Trp	Asp	Asn	Val	Leu	Asp	Val		
		195				200					205						
gca	cgc	gaa	aac	gcc	cag	gcg	gcc	aaa	gtg	gcc	gag	cga	gcg	cgt	ttc	672	
Ala	Arg	Glu	Asn	Ala	Gln	Ala	Ala	Lys	Val	Ala	Glu	Arg	Ala	Arg	Phe		
		210			215					220							
ctg	ccc	ggc	aac	gca	ttc	gac	ctc	gat	tac	ggc	agc	ggc	tac	gac	gtg	720	
Leu	Pro	Gly	Asn	Ala	Phe	Asp	Leu	Asp	Tyr	Gly	Ser	Gly	Tyr	Asp	Val		
225				230					235				240				
atc	ttg	ttg	acc	aac	ttc	ctg	cac	cat	ttc	gat	gag	gtc	gat	ggc	gag	768	
Ile	Leu	Leu	Thr	Asn	Phe	Leu	His	His	Phe	Asp	Glu	Val	Asp	Gly	Glu		
			245				250						255				
cgc	atc	ttg	gct	aag	acg	cgc	gat	gcg	ctg	aac	gac	gac	ggc	atg	gtg	816	
Arg	Ile	Leu	Ala	Lys	Thr	Arg	Asp	Ala	Leu	Asn	Asp	Asp	Gly	Met	Val		
		260				265					270						
atc	act	ttc	gaa	ttc	atc	gcc	gac	gaa	gag	cgt	tcc	tca	ccg	ccg	ctg	864	
Ile	Thr	Phe	Glu	Phe	Ile	Ala	Asp	Glu	Glu	Arg	Ser	Ser	Pro	Pro	Leu		
		275				280					285						

gcc gcc acc ttc agc atg atg atg ctg gcc acc acc ccg gcg ggc gag 912
 Ala Ala Thr Phe Ser Met Met Met Leu Gly Thr Thr Pro Ala Gly Glu
 290 295 300
 tcc tac acc tat agc gat ctg gaa agg atg ttt cgg cat gcc ggc ttc 960
 Ser Tyr Thr Tyr Ser Asp Leu Glu Arg Met Phe Arg His Ala Gly Phe
 305 310 315 320
 ggc cac gtg gaa cta aaa tcg ata ccg ccg gcc ttg ctg aaa gtg gtg 1008
 Gly His Val Glu Leu Lys Ser Ile Pro Pro Ala Leu Leu Lys Val Val
 325 330 335
 gtt tcc cgc aag agg gcc cca taa 1032
 Val Ser Arg Lys Arg Ala Pro
 340

<210> 97
 <211> 343
 <212> PRT
 <213> *Xanthomonas albilineans*

<400> 97

Met Asp Ser Ala Leu Pro Thr Ser Ala Phe Thr Phe Asp Leu Phe Tyr
 1 5 10 15

Thr Thr Val Asn Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile
 20 25 30

Glu Leu Gly Leu Phe Asp Val Val Gly Gln Gln Gly Arg Thr Pro Ala
 35 40 45

Ala Ile Ala Glu Ala Cys Gln Ala Ser Pro Arg Gly Ile Arg Ile Leu
 50 55 60

Cys Tyr Tyr Leu Val Ser Ile Gly Phe Leu Arg Arg Asn Gly Gly Leu
 65 70 75 80

Phe Tyr Ile Asp Arg Asn Met Ala Met Tyr Leu Asp Arg Ser Ser Pro
 85 90 95

Gly Tyr Leu Gly Gly Ser Ile Lys Phe Leu Leu Ser Pro Tyr Ile Met
 100 105 110

Ser Ala Phe Thr Asp Leu Thr Ala Val Val Arg Thr Gly Lys Ile Asn
 115 120 125

Leu Ala Gln Asp Gly Val Val Ala Pro Asp His Pro Gln Trp Val Glu
 130 135 140

Phe Ala Arg Ala Met Ala Pro Met Met Ala Leu Pro Ser Ala Leu Ile
 145 150 155 160

Ala Asn Met Val Ser Leu Pro Ala Asp Arg Pro Ile Arg Val Leu Asp
 165 170 175

Val Ala Ala Gly His Gly Leu Phe Gly Ile Ala Phe Ala Gln Arg Phe
 180 185 190

Arg Gln Ala Glu Val Ser Phe Leu Asp Trp Asp Asn Val Leu Asp Val
 195 200 205

Ala Arg Glu Asn Ala Gln Ala Ala Lys Val Ala Glu Arg Ala Arg Phe
 210 215 220

Leu Pro Gly Asn Ala Phe Asp Leu Asp Tyr Gly Ser Gly Tyr Asp Val
 225 230 235 240

Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Val Asp Gly Glu
 245 250 255

Arg Ile Leu Ala Lys Thr Arg Asp Ala Leu Asn Asp Asp Gly Met Val
 260 265 270

Ile Thr Phe Glu Phe Ile Ala Asp Glu Glu Arg Ser Ser Pro Pro Leu
 275 280 285

Ala Ala Thr Phe Ser Met Met Met Leu Gly Thr Thr Pro Ala Gly Glu
 290 295 300

Ser Tyr Thr Tyr Ser Asp Leu Glu Arg Met Phe Arg His Ala Gly Phe
 305 310 315 320

Gly His Val Glu Leu Lys Ser Ile Pro Pro Ala Leu Leu Lys Val Val
 325 330 335

Val Ser Arg Lys Arg Ala Pro
 340

<210> 98
 <211> 21
 <212> DNA
 <213> Xanthomonas albilineans

<220>
 <221> CDS
 <222> (1)..(21)
 <223> Motif I

<400> 98
 gtg ctg gac gtg gca gcc ggc
 Val Leu Asp Val Ala Ala Gly
 1 5

<210> 99
<211> 7
<212> PRT
<213> Xanthomonas albilineans
<400> 99

Val Leu Asp Val Ala Ala Gly
1 5

<210> 100
<211> 24
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(24)
<223> Motif II

<400> 100
agc ggc tac gac gtg atc ttg ttg
Ser Gly Tyr Asp Val Ile Leu Leu
1 5

24

<210> 101
<211> 8
<212> PRT
<213> Xanthomonas albilineans

<400> 101

Ser Gly Tyr Asp Val Ile Leu Leu
1 5

<210> 102
<211> 27
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(27)
<223> Motif III

<400> 102
ctg aac gac gac ggc atg gtg atc act
Leu Asn Asp Asp Gly Met Val Ile Thr
1 5

27

<210> 103
<211> 9
<212> PRT
<213> Xanthomonas albilineans

<400> 103

Leu Asn Asp Asp Gly Met Val Ile Thr
1 5

<210> 104
<211> 303
<212> DNA
<213> Xanthomonas albilineans

<220>
<221> CDS
<222> (1)..(303)
<223>

<400> 104
gtg ctg gac gtg gca gcc ggc cac ggc ctg ttc ggc atc gcc ttc gcg 48
Val Leu Asp Val Ala Ala Gly His Gly Leu Phe Gly Ile Ala Phe Ala
1 5 10 15
cag cgc ttc cgc cag gct gaa gtg agc ttc ctg gac tgg gac aac gtg 96
Gln Arg Phe Arg Gln Ala Glu Val Ser Phe Leu Asp Trp Asp Asn Val
20 25 30
cta gac gta gca cgc gaa aac gcc cag gcg gcc aaa gtg gcc gag cga 144
Leu Asp Val Ala Arg Glu Asn Ala Gln Ala Ala Lys Val Ala Glu Arg
35 40 45
gcg cgt ttc ctg ccc ggc aac gca ttc gac ctc gat tac ggc agc ggc 192
Ala Arg Phe Leu Pro Gly Asn Ala Phe Asp Leu Asp Tyr Gly Ser Gly
50 55 60
tac gac gtg atc ttg ttg acc aac ttc ctg cac cat ttc gat gag gtc 240
Tyr Asp Val Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Val
65 70 75 80
gat ggc gag cgc atc ttg gct aag acg cgc gat gcg ctg aac gac gac 288
Asp Gly Glu Arg Ile Leu Ala Lys Thr Arg Asp Ala Leu Asn Asp Asp
85 90 95
ggc atg gtg atc act 303
Gly Met Val Ile Thr
100

<210> 105
<211> 101
<212> PRT
<213> Xanthomonas albilineans

<400> 105
Val Leu Asp Val Ala Ala Gly His Gly Leu Phe Gly Ile Ala Phe Ala
1 5 10 15
Gln Arg Phe Arg Gln Ala Glu Val Ser Phe Leu Asp Trp Asp Asn Val
20 25 30
Leu Asp Val Ala Arg Glu Asn Ala Gln Ala Ala Lys Val Ala Glu Arg
35 40 45

Ala Arg Phe Leu Pro Gly Asn Ala Phe Asp Leu Asp Tyr Gly Ser Gly
50 55 60

Tyr Asp Val Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Val
65 70 75 80

Asp Gly Glu Arg Ile Leu Ala Lys Thr Arg Asp Ala Leu Asn Asp Asp
85 90 95

Gly Met Val Ile Thr
100

<210> 106

<211> 831

<212> DNA

<213> Xanthomonas albilineans

<220>

<221> CDS

<222> (1)..(831)

<223>

<400> 106

atg gat tca gcg tta cct aca tct gca ttt acc ttc gat ctc ttt tac 48
Met Asp Ser Ala Leu Pro Thr Ser Ala Phe Thr Phe Asp Leu Phe Tyr
1 5 10 15

acc acg gtt aac gcc tac tat cgc act gcc gca gtc aag gcg gcg atc 96
Thr Thr Val Asn Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile
20 25 30

gaa ctg ggg cta ttc gat gtg gtg ggg cag cag ggc cga act ccc gca 144
Glu Leu Gly Leu Phe Asp Val Val Gly Gln Gln Gly Arg Thr Pro Ala
35 40 45

gcc atc gcc gag gcc tgc cag gcg tcg ccg cgc ggc att cgc atc ctt 192
Ala Ile Ala Glu Ala Cys Gln Ala Ser Pro Arg Gly Ile Arg Ile Leu
50 55 60

tgc tat tac cta gta tcg atc ggt ttt cta cgc cgc aac ggt ggc ctg 240
Cys Tyr Tyr Leu Val Ser Ile Gly Phe Leu Arg Arg Asn Gly Gly Leu
65 70 75 80

ttc tac ata gat cgc aac atg gcc atg tac ctg gat cgt agt tcg ccc 288
Phe Tyr Ile Asp Arg Asn Met Ala Met Tyr Leu Asp Arg Ser Ser Pro
85 90 95

ggc tac ctg ggt ggc agc atc aag ttc ctg ctc tcg ccc tac atc atg 336
Gly Tyr Leu Gly Gly Ser Ile Lys Phe Leu Leu Ser Pro Tyr Ile Met
100 105 110

agc gcc ttc acc gat ctg acc gcc gta gtc agg acc ggc aag atc aac 384
Ser Ala Phe Thr Asp Leu Thr Ala Val Val Arg Thr Gly Lys Ile Asn
115 120 125

ctg gcg cag gac ggc gtg gtg gca ccg gat cac ccg cag tgg gtg gaa 432
Leu Ala Gln Asp Gly Val Val Ala Pro Asp His Pro Gln Trp Val Glu

130 135 140

ttt gca cgc gcg atg gca ccg atg atg gcg ctg ccc tcg gcg ttg atc 480
Phe Ala Arg Ala Met Ala Pro Met Met Ala Leu Pro Ser Ala Leu Ile
145 150 155 160

gcc aat atg gtg tcg ttg ccc gct gat cgg ccg att cgt gtg ctg gac 528
Ala Asn Met Val Ser Leu Pro Ala Asp Arg Pro Ile Arg Val Leu Asp
165 170 175

gtg gca gcc ggc cac ggc ctg ttc ggc atc gcc ttc gcg cag cgc ttc 576
Val Ala Ala Gly His Gly Leu Phe Gly Ile Ala Phe Ala Gln Arg Phe
180 185 190

cgc cag gct gaa gtg agc ttc ctg gac tgg gac aac gtg cta gac gta 624
Arg Gln Ala Glu Val Ser Phe Leu Asp Trp Asp Asn Val Leu Asp Val
195 200 205

gca cgc gaa aac gcc cag gcg gcc aaa gtg gcc gag cga gcg cgt ttc 672
Ala Arg Glu Asn Ala Gln Ala Ala Lys Val Ala Glu Arg Ala Arg Phe
210 215 220

ctg ccc ggc aac gca ttc gac ctc gat tac ggc agc ggc tac gac gtg 720
Leu Pro Gly Asn Ala Phe Asp Leu Asp Tyr Gly Ser Gly Tyr Asp Val
225 230 235 240

atc ttg ttg acc aac ttc ctg cac cat ttc gat gag gtc gat ggc gag 768
Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Val Asp Gly Glu
245 250 255

cgc atc ttg gct aag acg cgc gat gcg ctg aac gac gac ggc atg gtg 816
Arg Ile Leu Ala Lys Thr Arg Asp Ala Leu Asn Asp Asp Gly Met Val
260 265 270

atc act ttc gaa ttc 831
Ile Thr Phe Glu Phe
275

<210> 107
<211> 277
<212> PRT
<213> Xanthomonas albilineans

<400> 107

Met Asp Ser Ala Leu Pro Thr Ser Ala Phe Thr Phe Asp Leu Phe Tyr
1 5 10 15

Thr Thr Val Asn Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile
20 25 30

Glu Leu Gly Leu Phe Asp Val Val Gly Gln Gln Gly Arg Thr Pro Ala
35 40 45

Ala Ile Ala Glu Ala Cys Gln Ala Ser Pro Arg Gly Ile Arg Ile Leu
50 55 60

Cys Tyr Tyr Leu Val Ser Ile Gly Phe Leu Arg Arg Asn Gly Gly Leu
65 70 75 80

Phe Tyr Ile Asp Arg Asn Met Ala Met Tyr Leu Asp Arg Ser Ser Pro
85 90 95

Gly Tyr Leu Gly Gly Ser Ile Lys Phe Leu Leu Ser Pro Tyr Ile Met
100 105 110

Ser Ala Phe Thr Asp Leu Thr Ala Val Val Arg Thr Gly Lys Ile Asn
115 120 125

Leu Ala Gln Asp Gly Val Val Ala Pro Asp His Pro Gln Trp Val Glu
130 135 140

Phe Ala Arg Ala Met Ala Pro Met Met Ala Leu Pro Ser Ala Leu Ile
145 150 155 160

Ala Asn Met Val Ser Leu Pro Ala Asp Arg Pro Ile Arg Val Leu Asp
165 170 175

Val Ala Ala Gly His Gly Leu Phe Gly Ile Ala Phe Ala Gln Arg Phe
180 185 190

Arg Gln Ala Glu Val Ser Phe Leu Asp Trp Asp Asn Val Leu Asp Val
195 200 205

Ala Arg Glu Asn Ala Gln Ala Ala Lys Val Ala Glu Arg Ala Arg Phe
210 215 220

Leu Pro Gly Asn Ala Phe Asp Leu Asp Tyr Gly Ser Gly Tyr Asp Val
225 230 235 240

Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Val Asp Gly Glu
245 250 255

Arg Ile Leu Ala Lys Thr Arg Asp Ala Leu Asn Asp Asp Gly Met Val
260 265 270

Ile Thr Phe Glu Phe
275

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU01/01190

A. CLASSIFICATION OF SUBJECT MATTERInt. Cl. ⁷: C07K 14/195; C07H 21/04; C12N 15/52, 15/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN: File Reg, File CA (subsequence search of the individual sequences of Claim 1 combined with keywords ketoacyl reductase, polyketide, antibiotic, xanthomonas, albicidin in File CA), Index (CA, WPI, Medline, keyword xanthomonas albilineans, polyketide, gene)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	STN FILE MEDLINE ABSTRACT 2001060291 & G. HUANG <i>et al.</i> , Gene, 255(2), September 19 2000, pp. 327-333. See abstract and GENBANK sequences AF239749 and AF238750 and CAS Registry number 332004-68-9.	1-94
X	PUBMED ABSTRACT 10780924 & F. SCHAUWECKER <i>et al.</i> , Chem. Biol., April 2000, 7(4), pp. 287-297. See abstract and GenPept sequence AAF42473, positions 947-952	1-94
P,X	STN FILE MEDLINE ABSTRACT 20011526256 & G. HUANG <i>et al.</i> , Microbiology, March 2001, 147(3), pp. 631-642. See abstract.	1-94



Further documents are listed in the continuation of Box C



See patent family annex

* Special categories of cited documents:	
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
28 November 2001Date of mailing of the international search report
18 DEC 2001Name and mailing address of the ISA/AU
AUSTRALIAN PATENT OFFICE
PO BOX 200, WODEN ACT 2606, AUSTRALIA
E-mail address: pct@ipaaustralia.gov.au
Facsimile No. (02) 6285 3929Authorized officer

L.F. MCCAFFERY
Telephone No : (02) 6283 2573

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU01/01190

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	STN FILE MEDLINE ABSTRACT 2001087312 & G. HUANG <i>et al.</i> , <i>Gene</i> , 258(1-2), November 27 2000, pp. 193-199. See abstract and CAS Registry number 346735-79-3.	1-94
P,X	STN FILE MEDLINE ABSTRACT 2001122130 & G. HUANG <i>et al.</i> , <i>FEMS Microbiology Letters</i> , December 1 2000, pp. 129-136. See abstract.	1-94
X	STN FILE CA-ABSTRACT 132:103595 & K. MAYER <i>et al.</i> , <i>Nature</i> , 1999, 402(6763), pp. 769-777. See abstract and CAS Registry number 254869-45-9	1-94
X	STN FILE CA ABSTRACT 130:120325 & S. T. COLE <i>et al.</i> , <i>Nature</i> 1998, 396(6707), pp. 190-198. See abstract and CAS Registry numbers 208869-92-5 and 208869-94-7.	1-94
X	STN FILE CA ABSTRACT 123:331621 & F. BETSOU <i>et al.</i> , <i>Gene</i> , 1995, 162(1), pp. 165-166. See abstract and CAS Registry number 170560-61-9.	1-94
X	STN FILE CA ABSTRACT 112:173135 & P. GLASER <i>et al.</i> , <i>Mol. Microbiol.</i> , 1988, 2(1), pp. 19-30. See abstract and CAS Registry number 126469-81-6.	1-94
X	STN 111:72007 & P. GLASER <i>et al.</i> , <i>EMBO J.</i> , 1988, 7(12), pp. 3997-4004. See abstract and CAS Registry 121889-91-6.	1-94

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU01/01190

Box I Observations where certain claims were found unsearchable (Continuation of Item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos :
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos : 1-94 (all in part)
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
See attached sheet.

3. ☐ Claims Nos :
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II Observations where unity of invention is lacking (Continuation of Item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU01/01190

Supplemental Box

(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Box No: I(2)

A subsequence search in the Registry file of the peptides defined by Claim 1 resulted in 2608 sequences that contained the defined sequences. This corresponded to 1381 Chemical Abstracts. Furthermore, this result does not take into account that the claims include variants and biologically active fragments of each of the individual peptides. Accordingly it is not economical to search this result.

Moreover, the invention appears to lie in the identification of new domains that are involved in Albicidin synthesis in *X. albilineans*. However, the domains defined in the claims constitute as few as four amino acids. Whilst this may be the active site of the domains in question, the function of a domain will be dependent on the constitution and topology of an entire region of the protein. Accordingly, the present search has been limited to the domains of the multifunctional polyketide-peptide synthase gene that appear to possess the functionalities defined in Claim 1, namely positions 1230-3116 (acyl-CoA ligase region); 3423-4724 and 9117-10367 (ketosynthase regions); 6660-7142 (ketoreductase region); 3117-3422, 8598-8795 and 8859-9068 (acyl carrier regions); 12447-14066 (adenylation region); 10890-11150 and 14067-14306 (peptidyl carrier regions); 11151-12446 and 14307-15632 (condensation regions). Similarly the search of the PPTase motifs have been limited to the specific peptides defined by SEQ ID NO 89, 93, 91, 87, 99, 101, 103, 105, 107 and 91. It is not possible and/or economically viable to search for variants and biologically active fragments of these sequences.